Elucidating the molecular mechanisms of streptococcal M protein-induced autoimmune sequelae

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Giuseppe Gulotta
aus Mazara del Vallo / Italien

1. Referent: PD Dr. Daniel Patric Nitsche-Schmitz
2. Referent: Prof. Dr. Michael Steinert
eingereicht am: 15.12.2014
mündliche Prüfung (Disputation) am: 24.02.2015
Druckjahr 2015

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Tagungsbeiträge

D. Patric Nitsche-Schmitz, Giuseppe Gulotta, René Bergmann, G. Singh Chhatwal. The trigger of collagen IV autoimmunity in rheumatic fever. (Oral presentation). XIX Lancefield International Symposium on streptococci, Buenos Aires, Argentina. (2014).

Giuseppe Gulotta, René Bergmann, G. Singh Chhatwal & D. Patric Nitsche-Schmitz. Can PARF (peptide associated with rheumatic fever) be used to combat rheumatic fever? (Poster). XIX Lancefield International Symposium on streptococci, Buenos Aires, Argentina. (2014).

Giuseppe Gulotta, René Bergmann, G. Singh Chhatwal & D. Patric Nitsche-Schmitz. Collagen IV autoimmunity in rheumatic fever is triggered specifically by the PARF motif of streptococcal M proteins (Poster). 65th DGHM & DGI conference, Rostock,Germany. (2013).

Summary

Streptococcus pyogenes infects more than 600 million humans every year. In addition to a variety of acute infections, this bacterium causes acute rheumatic fever (ARF), an autoimmune sequela that frequently develops into rheumatic heart disease (RHD). RHD is a leading health problem in several low-resource regions in the world, thus remains a major public health concern. Evidence is crystallizing that *Streptococcus dysgalactiae* subsp. *equisimilis* is also a cause for ARF and RHD.

A vast variety of *S. pyogenes* serotypes and corresponding genotypes are distinguished based on variations in M protein or the M protein coding *emm* gene, respectively. M proteins are potent virulence factors and a major target of the human antibody response against this bacterium. In response to the immunity, the portion of the *emm* gene that encodes the N-terminal sequence of the M protein is under strong diversifying selection pressure and currently more than 200 distinct *emm*-types have been identified. Moreover, the M protein has been implicated in the pathogenesis of ARF. This protein triggers autoimmunity by molecular mimicry of host proteins with coiled-coil structure such as myosin or laminin. In addition, a collagen-binding A/T/E)xYLxx(L/F)N motif occurs in the N-terminal type specific part of M proteins that evoke collagen IV autoimmunity. This immune response against collagen IV is observed in ARF patients, but is not a cross reaction between M protein and the host protein as in molecular mimicry. As these observations suggested a link between the collagen binding motif and the autoimmune responses in ARF, the motif is referred to as "peptide associated with rheumatic fever" (PARF). However, the role of collagen-binding and of other properties of M proteins in triggering autoimmunity against collagen IV and consequences of this host response remain elusive.

Therefore, full-length M proteins of different *emm*-types were tested in this work for their ability to bind collagens I or IV and to evoke autoimmune responses against these collagens and the coiled-coil proteins myosin and laminin in a mouse model. The work proved for the first time that induction of collagen IV autoimmunity is PARF-specific, thus not caused by other parts or properties of the M protein. Furthermore, sufficient affinity of PARF carrying M proteins for collagen IV appeared to be crucial for triggering the aforementioned autoimmunity.

Further evidence for the role of PARF as a trigger of autoimmunity came from experiments with recombinant protein that contained all known naturally occuring PARF motifs (PARF tandem). Other than a synthetic PARF peptide (peptide 17) that was optimized for collagen binding, the multivalent PARF tandem induced autoimmunity against collagen IV and myosin in mice.

Due to the potential of PARF-positive M proteins to trigger autoimmunity, the corresponding strains may be important targets for vaccines. Notably, PARF tandem induced the production of opsonizing antibodies in rabbit, that reacted with PARF-positive M proteins of two distinct types. As immunization with PARF tandem caused autoimmunity too, a vaccine strategy that targets PARF would require further development.

Collagen-binding PARF occurs in about 7% of the *S. pyogenes* isolates that infect human beings worldwide. Moreover, the discovery of Spa as a novel collagen binding protein that triggers autoimmunity against collagen, despite lacking a PARF motif, broadens the spectrum of potential rheumatogenic factors in *S. pyogenes*. For the first time, a trigger of autoimmunity was identified in an *emm*-type (*emm*18) that is clearly associated with several outbreaks of ARF in the USA.

The presented work clearly points towards a role of interactions between host and microbial proteins in triggering autoimmunity and may have uncovered a new facet in the pathogenesis of autoimmune diseases in general. However, more research is needed for a comprehensive understanding of the underlying processes and their medical significance. Currently, we may only see the tip of an iceberg.

Zusammenfassung

Streptococcus pyogenes infiziert mehr als 600 Millionen Menschen im Jahr. Außer einer Vielfalt von akuten Infektionen, verursacht das Bakterium akutes rheumatisches Fieber (ARF), eine Autoimmunerkrankung, die häufig zur rheumatischen Herzerkrankung führt (RHD). RHD ist ein führendes Gesundheitsproblem in einigen Schwellen- und Entwicklungsländern. Es mehren sich die Hinweise, dass auch *Streptococcus dysgalactiae* subsp. *equisimilis* eine Ursache für ARF und RHD ist.

Eine enorme Vielfalt an *S. pyogenes* Serotypen und zugehörigen Genotypen werden auf Grund von Variationen im M Protein bzw. dem M-Protein-kodierenden *emm*-Gen unterschieden. Das M Protein ist ein sehr wirksamer Virulenzfaktor und Hauptangriffspunkt für die spezifische Immunantwort des Menschen. Letztere übt einen starken diversifizierenden Selektionsdruck auf den Teil des *emm*-Gens aus, der für den N-terminus kodiert, weshalb mehr as 200 verschiedene *emm*-Typen bekannt sind. Darüber hinaus wurde das M protein mit der Pathogenese von ARF im Zusammenhang gebracht. Das Protein löst Autoimmunität durch molekulare Mimikry von Wirtsproteinen mit Coiled-coil-Struktur aus wie Myosin und Laminin. Außerdem kommt in einigen N-terminalen typspezifischen Enden von M proteinen, die eine Autoimmunität gegen Kollagen IV auslösen, das kollagenbindende A/T/E)xYLxx(L/F)N-Motiv vor. Diese Immunantwort gegen Kollagen IV wird auch in Patienten mit ARF beobachtet. Sie ist aber keine Kreuzreaktion zwischen dem Wirtsprotein und dem M protein wie im Falle der molekularen Mimikry. Weil diese Beobachtungen einen Zusammenhang zwischen dem Kollagenbindungsmotiv und Autoimmunantworten in ARF nahelegen, wird das Motiv "peptide associated with rheumatic fever" (PARF) genannt. Die Rolle der Kollagenbindung und von anderen Eigenschaften des M Proteins im Auslösen der Autoimmunität gegen Kollagen IV und die Folgen dieser Wirtsantwort sind bisher nicht ausreichend untersucht.

Deshalb wurden in dieser Arbeit M proteine von verschiedenen *emm*-Typen auf ihre Fähigkeit getestet Kollagen I oder IV zu binden und Autoimmunantworten gegen diese Kollagene und die Coiled-coil Proteine Myosin oder Laminin auszulösen. Diese Arbeit wies zum ersten Mal nach, dass die Induktion der Kollagen-IV-Autoimmunität PARF-spezifisch ist und folglich nicht durch andere Teile oder Eigenschaften des M Proteins verursacht wird. Darüber hinaus schien eine ausreichende Affinität von PARF-positiven M Proteinen für Kollagen IV unabdingbar für das Auslösen der oben genannten Autoimmunität zu sein.

Weitere Beweise für die Rolle von PARF als Auslöser für Autoimmunität kamen von Experimenten mit einem rekombinanten Protein, das alle bisher bekannten natürlich vorkommenden PARF-Motive enthielt (PARF tandem). Anders als ein synthetisches PARF-Peptid (Peptid 17), das auf Kollagenbindung hin optimiert wurde, induzierte PARF tandem Autoimmunität gegen Kollagen und Myosin in Mäusen.

Wegen des Potentials PARF-positiver M Proteine Autoimmunität auszulösen, könnten die zugehörigen Stämme wichtige Ziele für Impfstoffe sein. Bemerkenswert war, dass PARF tandem die Produktion opsonisierender Antikörper im Kaninchen induzierte, die gegen zwei unterschiedliche M proteine reagierten. Da die Immunisierung mit PARF tandem aber auch Autoimmunität verursachte, müsste eine Impfstrategie, die auf PARF zielt, weiterentwickelt werden.

Kollagenbindendes PARF kommt in etwa 7% der *S. pyogenes*-Isolate vor, die den Menschen infizieren. Die Entdeckung von Spa als neues kollagenbindendes Protein, das Autoimmunität gegen Kollagen auslöst, obwohl es kein PARF-Motiv trägt, erweitert das Spektrum potentieller rheumatogener Faktoren in *S. pyogenes*. Zum ersten Mal wurde ein Auslöser von Autoimmunität in einem emm-Typ (*emm*18) entdeckt, der nachweislich mit mehreren Ausbrüchen von ARF und RHD in den USA assoziiert ist.

Die vorliegende Arbeit weist klar auf eine Rolle von Interaktionen zwischen Wirts- und Mikrobiellen Proteinen in der Entstehung von Autoimmunität hin und könnte generell eine neue Facette in der Pathogenese von Autoimmunerkrankungen aufgedeckt haben. Es ist jedoch mehr Forschung erforderlich, um die zugrundeliegenden Prozesse und ihre medizinische Signifikanz umfassend zu verstehen. Wir sehen gegenwärtig vielleicht nur die Spitze eines Eisbergs.

INDEX

CHAPTER 2 Materials and Methods

List of abbreviations

A

ARF: acute rheumatic fever

B

BSA: bovine serum albumin

C

C4BP: C4b-binding protein

CCR: conserved C-repeat region

CDC: Centers for Disease Control and Prevention

Cpa: collagen-binding protein of group A streptococci

D

DI water: deionized water

DNA: deoxyribonucleic acid

F

Fc region: fragment crystallizable region

G

GAC: group A carbohydrate

GAGs: gylcosaminoglicans

GAS: group A streptococci

GCGS: group C and G streptococci

GlcNAc: N-acetyl-β-d-glucosamine

H

HLA: human leukocyte antigen

I

IgG: immunoglobulin G

IL: interleukin

L

LA: Luria-Bertani agar medium

LB: Luria-Bertani medium

M

mAbs: monoclonal antibodies

Mga: multi gene activator

MHC: major histocompatibility complex

MLST: multilocus sequence typing

Mrp: M-related protein

P

PARF: peptide associated with rheumatic fever

PBS: phosphate buffered saline

PCR: polymerase chain reaction

PSGN: poststreptococcal glomerulonephritis

R

RHD: rheumatic heart disease

S

SCPA: streptococcal cysteine proteinase A

SDS: sodium dodecyl sulphate

SDSE: *S. dysgalactiae* subsp. *equisimilis*

Slr: streptococcal leucine rich

Spa: streptococcal protective antigen

STSS: streptococcal toxic shock syndrome

T

TNF: tumor necrosis factor

CHAPTER 1

General Introduction

1.1 The genus *Streptococcus*

The genus *Streptococcus* comprises species of non-motile Gram-positive extracellular bacteria. Streptococci are lactic acid fermenting bacteria and many of them colonize or infect humans or animals. The name *Streptococcus*, from the Greek (*streptos* = twisted chain) and (*coccus* = spherical), refers to the characteristic way in which these spherical bacteria grow in chains, like pearls on a string (Fig. 1).

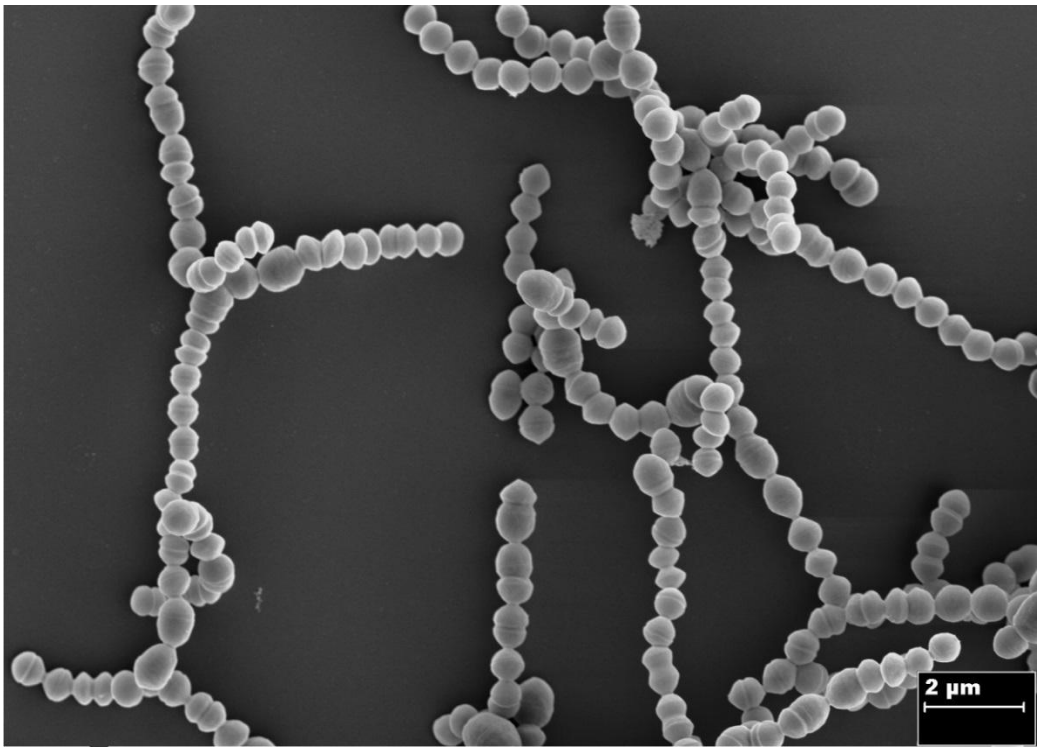


Figure 1: **Chains of *S. pyogenes* at 5000 fold magnification.** The picture was kindly provided by Prof. M. Rohde (HZI-Braunschweig).

The concept of species is not well defined in bacteriology and faces its limits when studying bacterial genomes, because of lateral gene transfer. During bacterial evolution, lateral gene transfer is responsible of the movement of genetic material even between distantly related bacteria, blurring the borders between taxonomic entities (**1**). However, grouping of bacteria based on metabolic properties or their genome correlates with their ecology and, not least,

their pathogenic potential. Hence, dividing bacteria into species is of great value in research, biotechnology and medicine.

The genus *Streptococcus* includes more than 50 genetically distinct species and subspecies (**2-3**), which are divided into six groups: the mitis group, the anginosus group, the salivarius group, the mutans group, the bovis group and the pyogenic group (Fig. 2). Streptococcal species differ substantially in their ecology and their pathogenic potential in humans or animals. The pyogenic group for instance contains prominent human pathogenic streptococci such as *Streptococcus pyogenes*, *Streptococcus agalactiae* and *Streptococcus dysgalactiae* subsp. *equisimilis* (1.3).

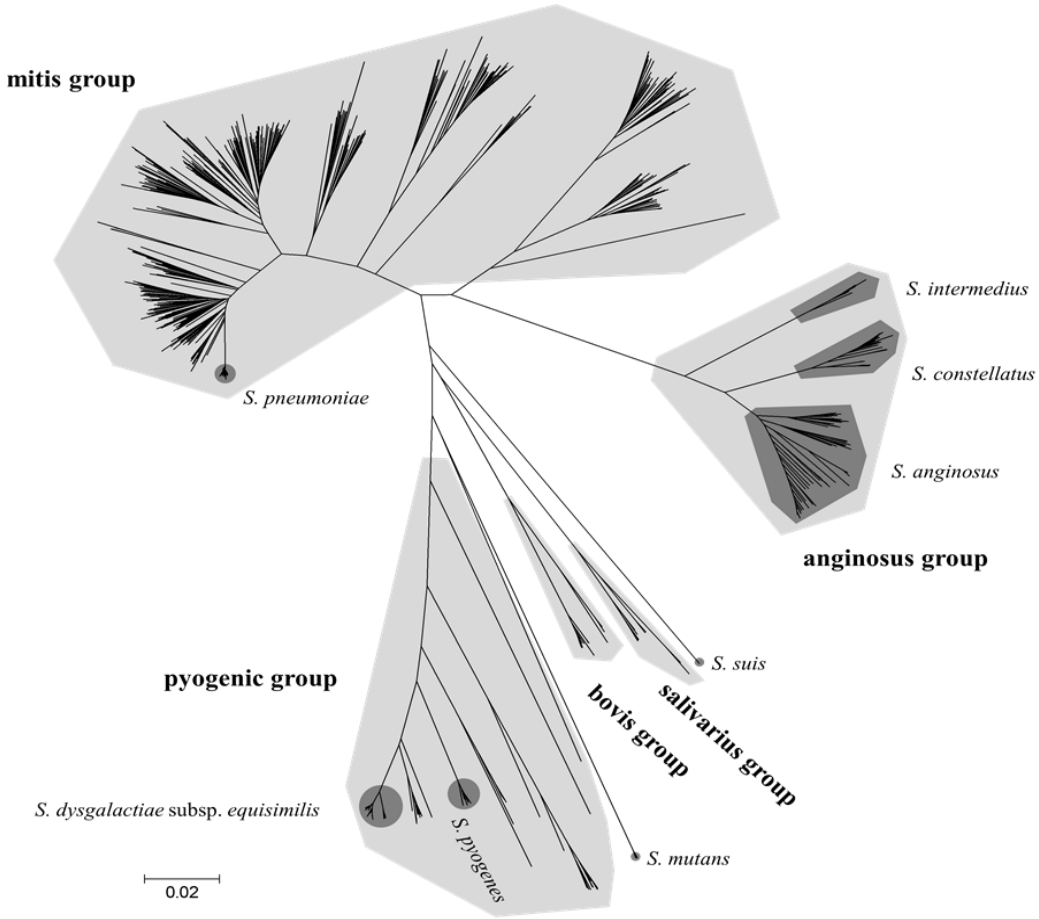


Figure 2: **Neighbor joining tree of a multilocus sequence analysis of 852 streptococcal isolates of different (sub-) species.** Streptococcal groups and selected (sub-) species are indicated in light grey and dark grey, respectively. The analysis comprises data of different studies (**4-5-6**)

1.2 Distinguishing streptococci

1.2.1 Hemolysis

A historically early method to distinguish streptococci is based on their ability to lyse erythrocytes on blood agar plates and to degrade the hemoglobin, a phenomenon called hemolysis. According to the degree of hemolysis, these bacteria are divided in α, β and γ-hemolytic streptococci (**7**). α-hemolytic streptococci oxidize the iron ion of the hem and partially degrade this prosthetic group of the hemoglobin, causing incomplete hemolysis. This appears as a greenish halo that surrounds the bacterial colonies on blood agar. β-hemolytic species cause complete rupture of the erythrocytes and degrade the hem completely. On blood agar plates this appears as a clear zone that surrounds the bacterial colonies. γ-hemolytic streptococci do not cause hemolysis. The type of hemolysis is a valuable diagnostic parameter, as important pathogenic streptococci are β-hemolytic (**8**).

1.2.2 Lancefield-grouping

Streptococci are further distinguished by a serological method that has been established by Rebecca Lancefield in 1932 (**9**). This method is based on group-specific antigens on the bacterial surface called Lancefield group antigens. Lancefield-grouping separates streptococci into more than 20 distinct serological groups, which were designated with Latin letters starting from A. This method is widely used in diagnosis, as it identifies two prominent human pathogenic streptococcal species, *S. pyogenes* (group A) and *S. agalactiae* (group B). However, this method fails to distinguish between different taxonomic entities of other Lancefield groups, for instance streptococci of the Lancefield groups C and G (GCGS) (1.4 SDSE). GCGS comprise several species with different pathogenic potential in humans or animals (**10**). Consequently, Lancefield-grouping is of certain but limited value in the characterization of streptococci and diagnosis of streptococcal infections.

1.2.3 DNA based methods

As mentioned above, the concept of species is not well defined in bacteriology and consequently, there is no clear-cut definition for bacterial species. More or less arbitrary thresholds of DNA homology are used when distinguishing streptococcal species (**11**). The percentage of DNA sequence homology is determined by DNA-DNA hybridization or whole genome sequencing (**11-12**). Other than DNA-DNA hybridization, whole genome sequencing delivers additional information about the gene content or genetic rearrangements in the investigated bacteria. Selected members of the genus are described below.

1.3 Viridans streptococci and the bovis group

The name viridans streptococci, derives from "*viridis*", the Latin word for green. It is a non-taxonomic term for streptococci that produce a greenish coloration on blood agar plates, due to α-hemolysis (1.2 Hemolysis). While the majority of the viridans streptococci is α-hemolytic, some strains of these bacteria are β- or γ-hemolytic and show other phenotypic inconsistencies even within species (**9**). Most of the viridans streptococci are commensals of the human oral cavity, which are referred to as oral streptococci. Although they are commensals in certain niches of the human host, viridans streptococci are facultative pathogens (**6**). They comprise four out of the six groups of streptococci: the anginosus group, the mitis group, the salivarius group and the mutans group (**6**).

The anginosus group includes streptococci that can be part of the normal flora of the oral cavity, the urogenital tract and the gastrointestinal tract of humans. Nonetheless, streptococci from the anginosus group are facultative human pathogens (**5**). Infections with these streptococci can have a pyogenic (pus-forming) character and some strains share typical properties with pyogenic streptococci like β-hemolysis and Lancefield antigens A, C or G. Other strains are not groupable or belong to group F (**13**). The spectrum of diseases that is

caused by the anginosus group comprises severe infections in the oral cavity, the skin, the central nervous system and in inner organs like the brain or the liver. (**5**) Streptococci of this group have a characteristic and pronounced tendency to cause abscesses and their infections can become life-threatening (**14**).

The mutans group is named after the prototypic bacteria of this group, *S. mutans*, which is a part of the dental plaque and the leading cause for dental caries in humans (**15**).

Streptococci of the mitis and of the salivarius group colonize the human oral cavity where they are involved in caries formation and gingivitis (**16**). Moreover, they can invade the blood stream, occasionally causing life-threatening diseases (**17**). Streptococci of the salivarius group are a frequent cause of septicemia in neutropenic patients (**18**) while streptococci of the mitis group are the leading causes of streptococcal endocarditis (**6-19**). This is an infection of the endocardium that usually involves the heart valves and has high morbidity and mortality world-wide (**6**). An important predisposing factor for infective endocarditis is acute rheumatic fever (**20**) (1.7.2 ARF and RHD).

The bovis group includes streptococci that are inhabitants of the human gastrointestinal tract (**20-21**). These streptococci cause bacteraemia and endocarditis, often associated with perforating gastrointestinal diseases (**22**) or colorectal cancer (**23-24**). A potential role of these bacteria in the development of cancer is under investigation. It is not yet clear whether bovis group streptococci are causative for gastrointestinal malignancy or whether malignancy facilitates colonization with these bacteria.

1.4 Pyogenic streptococci

The pyogenic group includes the most important β-hemolytic streptococcal pathogens that cause various suppurative (pus-producing) infections in humans or animals. Production of pus

is the result of the attraction and activation of leucocytes, at the site of infection. As a defence against these cells of the immune system, streptococci produce toxins that kill the leucocytes, contributing to the pyogenic character of the infection. A well-known human pathogen of the pyogenic group is *S. agalactiae*, which forms the Lancefield group B (1.2 Lancefield-grouping). This bacterium is part of the normal flora of the gastrointestinal and genital tract of humans and animals. As a colonizer of the female genital tract, it poses a significant obstetric problem. *S. agalactiae* causes puerperal fever and is the leading cause of bacterial septicemia and meningitis in newborns children (**25**). Other important human pathogens of the pyogenic group are *S. pyogenes* and *S. dysgalactiae* subsp. *equisimilis* (SDSE). Worldwide, they are responsible for more than 800 million infections every year (**26**).

1.4.1 *Streptococcus pyogenes*

S. pyogenes is a β-hemolytic *Streptococcus* of the pyogenic group and belongs to the Lancefield group A. This human specific pathogen is occasionally found in the flora of the upper respiratory tract of symptom free individuals (**27**). To date, humans are the only known reservoir of these bacteria, which are responsible for a wide spectrum of acute infections and cause severe post-infection sequelae (**28**). Worldwide *S. pyogenes* is the leading cause of bacterial pharyngitis (strep throat) and causes localized skin infections (pyoderma). If deeper tissues become infected with this pathogen, the production of bacterial toxins in combination with overreacting immune cells can lead to severe soft tissue infections such as necrotizing fasciitis. Necrotizing fasciitis is a life-threatening condition, often requiring surgical debridement of infected tissue or amputation. Even intensively treated necrotizing fasciitis can lead to death by streptococcal toxic shock syndrome (STSS) (**29**). Streptococcal toxins are also the cause of the characteristic rush that appears in scarlet fever, a disease that commonly affects children (**30**). Furthermore, acute streptococcal infections can trigger

deleterious immune responses that lead to the post-streptococcal sequelae acute rheumatic fever (ARF) and post-streptococcal glomerulonephritis (1.7.1).

History is replete with descriptions of epidemics caused by *S. pyogenes*, highlighting the extraordinary virulence of this bacterium. Well-documented examples are the outbreaks of scarlet fever that occurred in the latter part of the 1880s in New York, Chicago and Norway and in which 25% to 30% of the infected children died (**31-32**). Improvement of socio-economic conditions and the use of antibiotics during the last seven decades substantially reduced the incidence of infections caused by this pathogen in developed economic countries. Luckily, *S. pyogenes* has not yet developed resistance to β-lactam antibiotics, which remain the drugs of choice for treatment of acute streptococcal infections and for prevention of their sequelae (**33**). However, resistance of *S. pyogenes* to other antibiotics is on the rise. For instance, rates of resistance against erythromycin are high in several regions of the world (**34-35**). Despite the availability of effective antibiotics, the incidence of streptococcal acute infections and prevalence of their sequelae remains high, accounting for more than half a million deaths per year (**26**). Failure of antibiotic treatment in individual cases and in further reducing the burden of *S. pyogenes* infections and sequelae creates a need for novel therapeutic approaches, including a vaccine against this bacterium.

1.4.2 *Streptococcus dysgalactiae* subsp. *equisimilis*

Streptococci of Lancefield group C and G (GCGS) are widely distributed in both animals and humans. They can be isolated from the normal flora of the pharynx, the skin, the gastrointestinal tract and the female genital tract, where they can also cause infections (**36**). GCGS can be divided into the species *S. dysgalactiae*, *S. equi*, *S. canis* and *S. phocae*. Moreover, some strains of anginosus group streptococci can carry the group C or G antigen (**36**).

For a long time, GCGS were considered mainly as animal pathogens. Moreover, taxonomic heterogeneity of GCGS led to difficulties in the diagnosis of infections caused by these pathogens. Therefore, the pathogenic potential that GCGS have in humans has long been underestimated. Today, due to improvements in diagnostic methods and increasing awareness, GCGS are recognized as relevant human pathogens (**37-38**).

One pathogen of the GCGS is *Streptococcus equi* subsp. *zooepidemicus.* This zoonotic pathogen causes septicemia, purulent arthritis and post-streptococcal glomerulonephritis (PSGN) in humans (**39**).

The dominant human pathogen of the GCGS is *Streptococcus dysgalactiae* subsp. *equisimilis* (SDSE) (**40**). Streptococci of this subspecies share many virulence factors with *S. pyogenes* (**40**), for instance the M protein (1.6.1), the C5a peptidase (**41-42**), the hyaluronic acid capsule (**43**), hemolysins (**44**) and the streptokinase (**45**). This is considered as the reason why SDSE infect humans in a way that closely resembles *S. pyogenes* infections. SDSE causes pharyngitis and suppurative skin and soft tissue infections (**46-47**).

Experimental studies indicate a potential role of SDSE in triggering ARF (**48-49**). This is supported by epidemiological studies in some high prevalence regions for ARF, in which SDSE was the prevailing species in streptococcal pharyngitis. For instance, more than 60% of the throat isolates from a population of Indigenous Australians belonged to the subspecies SDSE, exceeding the isolation rate of *S. pyogenes* (**50**). Moreover, SDSE caused 75% of the streptococcal pharyngitis cases that occurred in 1978 in Lagos, Nigeria (**51**).

The clinical importance of SDSE appears to be increasing (**52**). Recent epidemiological studies indicate an increase of invasive SDSE infections that often occur among elder or immunocompromised persons. Thus, it seems advisable to include these streptococci in health

care strategies that aim at reducing the burden of acute infections by pyogenic streptococci and their sequelae.

1.5 Typing of *S. pyogenes* and SDSE isolates

In order to investigate and describe the high variety of streptococcal isolates and to uncover correlations between certain types and certain ecological or pathogenic characteristics, researchers studied different serological and genetic markers of *S. pyogenes*.

1.5.1 M serotyping

S. pyogenes can be serologically divided into different M serotypes based on the streptococcal M antigen. The M antigen is a protein called M protein (1.6.1) and encoded by the *emm* gene. This surface bound protein differs between different isolates in its N-terminal sequence and elicits the production of specific antibodies in infected humans (**53**). Raised in animals, antibodies against M protein are exploited for discrimination of streptococcal isolates of more than 80 different serotypes (**28**). However, the vast variety of M proteins, difficulties in standardizing M serotyping, limited availability of standardized M antisera worldwide and ambiguities due to cross-reactivity in the identification of M types limit the value of this method for larger studies and epidemiological surveys. Therefore, M serotyping has been virtually replaced by the genetic *emm*-typing, a standardized method that is based on the sequence of the *emm* gene. It offers a higher resolution than M serotyping and is explained below.

1.5.2 T-typing

The T-typing assay, performed as an agglutination test, has been used in epidemiological investigations of *S. pyogenes* infections. The T antigen, coded by the *tee* gene, is an antigen

that was detected on the surface of more than 95% of *S. pyogenes* isolates (**54**). This method allowed typing *S. pyogenes* isolates of an outbreak, when the M serotype was not identifiable (**55**). Moreover, certain T serotypes are associated with specific M protein serotypes (**55-56**). This knowledge was used to preselect and reduce the M antisera required for M-typing, allowing a faster determination of the M type (**27**).

1.5.3 *Emm*-typing

The *emm*-typing, established by the Centers for Disease Control and Prevention (CDC), is a sequence based approach. For *emm*-typing, the 5'-sequence of the *emm* gene that encodes the hypervariable N-terminus of the M protein (1.6.1) is analysed and used to discriminate *emm*-types and *emm* subtypes based on homology (**57**). Accordingly, the *emm*-type correlates with the M-type. This allows a fast and standardized typing of streptococci that is widely used for epidemiological studies on *S. pyogenes*, but also on SDSE, which also harbour an *emm* gene (**58-59**, 1.5.2). The *emm* gene is a genotype marker of *S. pyogenes* and SDSE and enables to type these bacterial isolates more efficiently than serological methods. An on-line database of *emm* genes is available, facilitating the exchange of information between different laboratories. The *emm*-typing revealed the diversity of *S. pyogenes* isolates (**60**). Currently, more than 200 *emm*-types have been identified and more will be identified in the future. Regional *emm*-typing data describe the spectrum of the circulating genotypes in defined geographic locations, which is of great scientific interest and of great importance in the design of streptococcal vaccines (1.5).

1.5.4 Vir-typing

Vir-typing is based on the multiple genes regulator (*mga*) region of *S. pyogenes* (1.6.1 and Fig. 3). The *mga* region is a 4 to 7 kb sequence that consists of the transcriptional regulator

Mga, genes that code for M and M-like proteins *emm*, *enn*, *mrp/fcR* and the *scp*A gene that encodes for the C5a peptidase (**61**). After amplification of the *mga* region by PCR, HaeIII and HinfI endonucleases are used to digest the amplicons in order to generate a restriction fragment length polymorphism (RFLP) profile. The resulting pattern determines the Vir-type, which correlates with the M type (**62**).

1.5.5 MLST

Multilocus sequence typing (MLST) is a general and widely used procedure for genotypic characterization of bacterial isolates. MLST of *S. pyogenes* isolates uses internal sequences of ~ 500 base pairs of the seven housekeeping genes *gki*, *gtr*, *murI*, *mutS*, *recP*, *xpt*, and *yqiL* (**62-63**). According to allelic profiles, streptococcal isolates are assigned to a sequence type (ST). The advantage of MLST is that sequence data are unambiguous and the allelic profiles of isolates can easily be compared to those in a large central database via the Internet www.mlst.net (**64**). The high resolution of MLST is used to investigate evolutionary relationships among the bacteria and to track the development and spread of certain clonal complexes. Moreover, samples like clinical material can be analysed by PCR amplification in a culture independent manner. For these reasons, MLST is increasingly applied in research and clinical laboratories.

1.5.6 Genotype diversity of *S. pyogenes*

S. pyogenes isolates exhibit a high genetic diversity. The necessity to distinguish them for scientific and clinical purposes has driven the research towards several typing methods. In the past, serological methods were utilized to study the epidemiology of *S. pyogenes* isolates, in particular how their types correlate with certain manifestations of infections in humans. With the advent of DNA sequencing, serological methods have been replaced by genotyping, such

as the *emm*-typing. In addition, whole genome sequencing has shed light on the marked genetic differences that occur even between *S. pyogenes* bacteria of the same *emm*-type (**65**). In contrast to whole genome sequencing, the available typing methods are relatively inexpensive ways to investigate the genetic variety collections of streptococci and to study their distribution in defined geographic regions and/or over time. Among the several typing methods, *emm*-typing is the most widely used and shed light on large geographic differencies and a considerable dynamic of *S pyogenes* genotype distribution.

Geographic regions differ in the diversity of *emm*-types and in prevalence of certain *emm*-types. According to Smeesters *et al.*, three groups are observed. In regions of the first and the second group, low *emm*-type diversity with a few predominant *emm*-types is observed. North America and Western Europe belong to the first group, sharing a similar spectrum of *emm*-types and the same prevalent types. The second group includes China and countries from Eastern Europe in which the predominant *emm*-types are different from the ones of the first group. In contrast, high *emm*-type diversity is observed in regions of the third group. In addition, no predominant *emm*-types were observed in these regions, most of which are high prevalence regions for ARF and RHD. For instance, newly industrialized countries like Brazil and India belong to the third group (**66**).

Early studies tried to identify correlations between certain M or *emm*-types, and colonization sites or manifestation of disease like pharyngitis, skin infections or ARF (**67**). For instance, studies on endemic outbreaks of ARF in the USA indicated that the rheumatogenic potential of *S. pyogenes* is limited to few M types corresponding to *emm* M1, M3, M5, M6, M18, M19, and M24 (**68-69-70**). Further studies associated M3 serotype with ARF-like side effects and M18 was associated with ARF outbreaks in USA (**71-72**). Notably, these correlations could not be observed in other parts of the world where other M-types are prevalent and SDSE

infections are more frequent (**73-74-75**). Consequently, connections between *emm*-type and diseases remain somewhat uncertain.

Although, epidemiological data were not able to picture the association between certain M types and certain human diseases, the genetic diversity of M protein is believed to play an important role in the spectrum of pathologies caused by *S. pyogenes* due to its several functions i.e. adhesion to the host, prevention of opsonophagocytosis, molecular mimicry with several host proteins and due to its possible role in triggering ARF (**76**).

M and *emm*-typing shed light on serotype replacement of *S. pyogenes.* Serotype replacement has been observed over a period of 7 years in London, where the isolates of M-types 6, 49 and 81 were replaced by types 1, 3 and 28 (**77**). Moreover, serotype replacement occurred within a short time period in a semi-closed community in the USA. In this community, during a time frame of three months, serotype M6 has replaced serotype M1, most likely due to post-infectious protective responses against type M1 (**78**).

Studying the *emm*-type distribution *S. pyogenes* and SDSE on the global scale as well as their prevalence in defined regions has provided valuable information for vaccine development. A vaccine designed to protect against *emm*-types that are prevalent in defined geographic regions could reduce the impact of streptococcal disease on human health (1.8.2). However, serotype replacement may be a problem for the success of such a vaccine. Thus, *emm*-typing could be an important and necessary method also in the future to studying serotype replacement that is driven by immune pressure due to vaccination.

1.6 M and M-like proteins of pyogenic streptococci

1.6.1 M protein

The M protein is a surface protein and one of the most studied virulence factors of *S. pyogenes* (**79-80**) and SDSE **(81-82).** Due to hypervariability in the 5'-end of the M protein gene (*emm*) more than 200 distinct *emm*-types have evolved and have been identified to date (www2a.cdc.gov/ncidod/biotech/strepblast.asp, as in November 2014).

M proteins and the related M-like proteins (1.6.2) are encoded by the *emm*, *mrp* and *enn* genes. Those genes are tandemly arranged in the Mga region of the *S. pyogenes* genome and under positive control of the Mga (multi-gene activator) regulator (Fig. 3), which also regulates the expression of the genes *sof* and *sfbX* that are located in trans (Fig. 3, **83**). The *scpa* gene that also included in the Mga region encodes the surface bound C5a peptidase, a virulence factor that is not related to the M proteins (**83**). The patterns of the Mga region can vary in *S. pyogenes* isolates according to their preferential colonization sites in the host. Mga pattern A, B and C caused mainly infectections of the oral tissues, whereas strains with pattern D were mostly isolated from the skin. *S. pyogenes* with a Mga region of pattern E were isolated at both oral and skin sites (**84-85**).

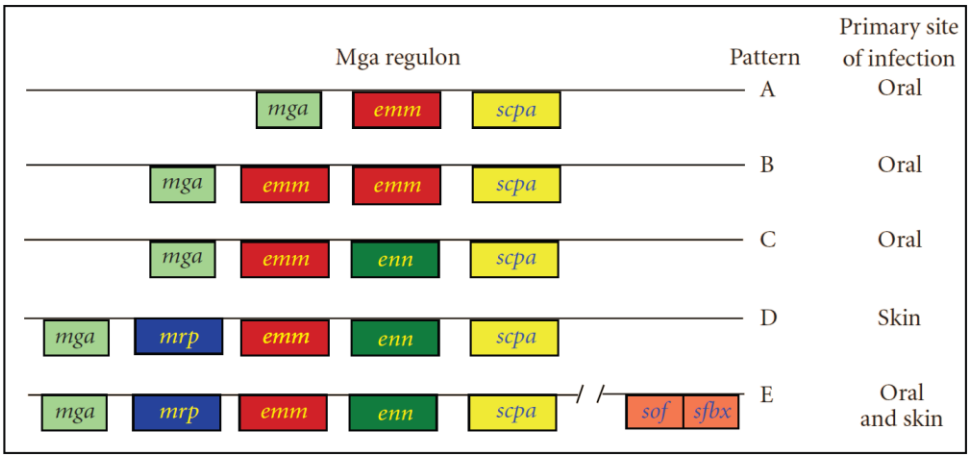


Figure 3: **Variants of the Mga region and their relationship to preferential colonization sites in the host.** The figure was taken from Courtney *et al.* (**86**).

1.6.1.1 Structure of the M protein

Electron microscopy revealed that the M proteins are arranged in a fur-like structure on the bacterial surface (Fig. 4). The M protein is a thread-like homodimer that forms a coiled-coil structure and extends ~50 nm from the streptococcal surface (**87-88,** Fig. 4).

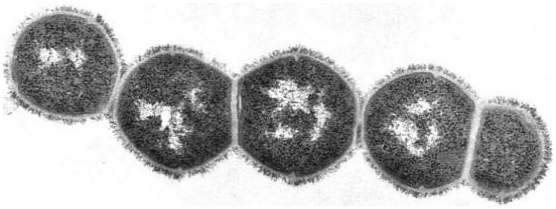


Figure 4: **Electron micrograph of an ultra-thin section of a chain of *S. pyogenes* group A streptococci (20.000 fold magnification)** The fur-like structure determined by the M proteins is clearly visible on the bacterial surface. (http://www.rockefeller.edu/vaf/chain20.php).

Each chain of the M protein dimer consists of up to four distinct repeat regions, named A, B, C and D starting from the N-terminus (**89**). Although the overall organisation of the M protein is conserved, major differences exist, not only in the variable N-terminus, but also in the number of the repeats (**76-90-91,** Fig. 5).

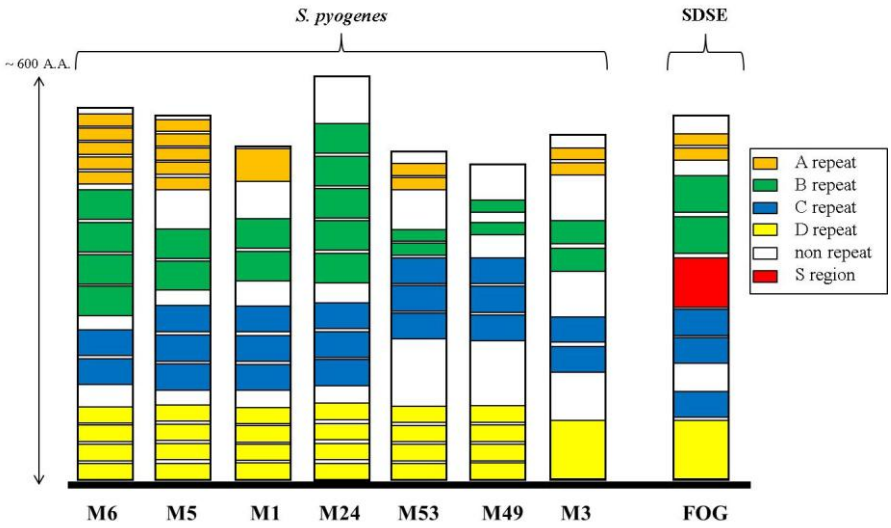


Figure 5: **Schematic representation of the mature form of eight distinct M proteins.** Presence, number and size of the repeats of the protein M6, M5, M1, M24, M53and M49 were deduced from (**75**). The composition of M3 and FOG were deduced from (**92-93**).

The common makeup of M protein monomers includes a conserved N-terminal signal peptide for extracellular secretion (**94**). The signal peptide sequence not only marks the molecule for export, but also targets secretion of the protein to the division septum (**95**). During maturation of the protein, the signal peptide is cleaved off by a signal peptidase and the M protein is covalently attached to the rigid peptidoglycan layer of the Gram-positive bacterial wall (Fig. 6) through the action of an enzyme that is called sortase (**96**). After cleavage of the signal peptide, the mature M protein has a hypervariable N-terminus. More central parts are less variable and the C-terminus is conserved (**76-79**).

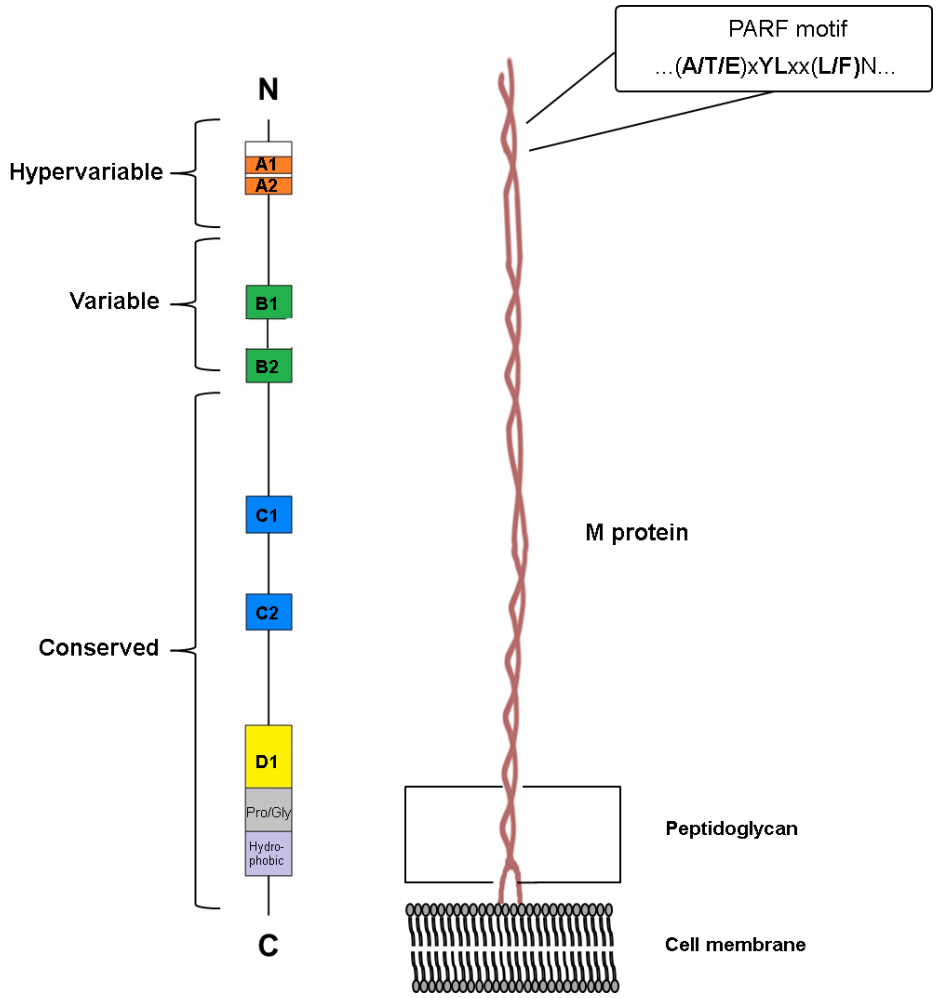


Figure 6: **Schematic structure of the M3 protein.** Blocks A, B, C, and D designate the location of the sequence repeat blocks. Pro/Gly denotes the proline-rich andglycine-rich region likely located in the peptidoglycan. The C-terminal end is located within the cell wall and membrane (**97**). The consensus sequence of the PARF motif is depicted in a box that also indicates the N-terminal position of PARF in M proteins (**59**).

Long stretches of the M protein sequence contain heptad repeats. Due to the α-helical structure, the first (a) and fourth (d) amino acid of the heptad repeat within an M protein chain are able to form hydrophobic bonds with their counterparts on the homologous M protein chain. The fifth (e) and seventh (g) amino acid form ionic interchain bonds (Fig. 7).

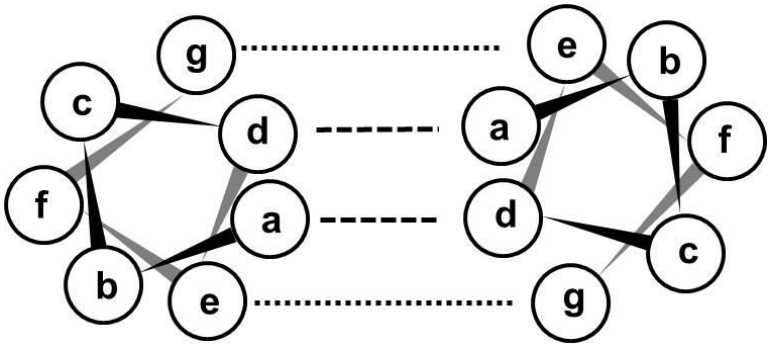


Figure 7: **Schematic representation of the interactions that occur in an M protein coiled-coil.** Position **a** and **d** are occupied by hydrophobic amino acids, **b**-**c**-**f** are polar amino acids, **e**-**g** are charged amino acids **.**

These interactions of core residues within the coiled-coil are important for the stability of the coiled-coil structure (**98**). The continuity of the heptad pattern is often disrupted because of insertion of additional amino acid residues, suggesting that M proteins have a flexible coiled-coil structure. The biological relevance of these irregularities is unknown, but they are also characteristic for coiled-coil proteins of eukaryotic origin, such as myosin and tropomyosin (**99**). Structural irregularities that occur in myosin and tropomyosin as well as in the M1 protein seem to be involved in evoking the production of cross-reactive antibodies by molecular mimicry. This may explain some of the autoantibody titers seen in patients with streptococcal autoimmune sequelae (**98**, 1.7.2 ARF and RHD).

The N-terminal hypervariability of the M protein results in antigenic variation (**28**-**100**) and is exploited for serotyping and, on the DNA level, for *emm*-typing (1.5 *Emm*-typing). Moreover, the N-terminal part of M protein elicits the production of type-specific protective

antibodies and is therefore considered as a valuable *S. pyogenes* vaccine candidate (1.8.2 M protein based vaccines).

Depending on the *emm*-type, the N-terminus of the M protein harbours the peptide associated with rheumatic fever (PARF) (Fig.6), a binding site for collagen IV. M proteins with a PARF motif triggered collagen IV autoimmune responses in mice, suggesting that the motif could take part in eliciting rheumatogenic processes (**59-101-102**). The role of the PARF-motif in inducing ARF-related autoimmune responses is further investigated in this work.

The hypervariable N-terminal sequence of M protein can contain A-repeats and is followed by the variable region that includes the B-repeats (Fig. 6). The B-repeats bind human fibrinogen, contributing to the antiphagocytic property of M protein (**103**). In some M proteins such as M1, M55 or stG11 the B-repeats are followed by the S-region, which binds human immunoglobulin G (IgG) in a non-immune fashion, via the Fc region of the IgG (**104-105-106**). Efficient bacterial IgGFc-binding by M protein occurs only in IgG-poor environments, such as saliva, and protects streptococci against phagocytic killing (**107**). In blood plasma, where the concentration of IgG is high, the antibodies bind M proteins mainly via Fab, which facilitates opsonization and bacterial killing. IgG-poor environments represent the natural habitat for IgGFc-binding bacteria and IgGFc-binding proteins may have evolved to exert their function in such environments. The lack of protection in plasma also helps to explain why cases of severe invasive infections with IgGFc-binding bacteria are so rare compared with more superficial and uncomplicated infections of the throat and skin (**106**).

The C-terminal region, well conserved between M proteins of different *emm*-types, is composed of the C- and D-repeats and the LPXTG-motif, which is followed by a sequence of wall-associated amino acids. This terminal wall-associated part of the M protein sequence is divided into two regions, one that is rich in proline and glycine. The other region is a

membrane anchor that consists of 19 hydrophobic amino acids (**79**). The M protein ends with a C-terminal sequence of six charged amino acids of unknown function. During the maturation of the M protein, a sortase catalyzes the cleavage between Thr and Gly of the LPXTG consensus motif and covalent binding to the streptococcal peptidoglycan wall (**107**). The LPXTG consensus motif is highly conserved in surface proteins of streptococci and other Gram-positive bacteria (**107**).

1.6.1.2 Functions of the M protein and its role in pathogenesis

The M protein is an adhesin, antiphagocytic factor and potent trigger of inflammation of *S. pyogenes* that contributes to the virulence and pathogenicity of this bacterium (**28**).

Adhesion of bacteria to dermal tissue or mucous membranes is an initial and crucial step in establishing streptococcal infection in the host. *S. pyogenes* expresses a panel of adhesion factors that bind strongly to host tissue components to counteract removal by protective host mechanisms such as the flow of saliva or blood. Different streptococcal genotypes can express an individual set of adhesins that interact with specific targets in the host. This depends on the strain's genetic makeup and its regulation of gene expression (**108**).

The C-terminus of the M protein binds several host molecules like the complement regulator factor H (**109**) and human serum albumin (**102-105-110-111**) promoting the virulence and the persistence of *S. pyogenes* in the host (**112**). Moreover, some M proteins exert adhesin properties by binding to several collagen types (**113**) or to gylcosaminoglicans (GAGs) (**114**). These interactions attach the streptococci to the extracellular matrix of the host or to cells such as human epithelial cells and skin fibroblasts (**115**).

M protein promotes the aggregation of *S. pyogenes*. The aggregation is mediated by a 19 amino acid long segment in the B-repeat region of M1 protein (**116**). This mechanism

increases resistance to phagocytosis and is involved in adherence to and invasion into epithelial cells, thereby contributing to the virulence of *S. pyogenes* (**116**).

Many M proteins interact with the complement regulator C4b-binding protein (C4BP), which inhibits the classical pathway of the complement system under non-immune conditions and may protect streptococci against phagocytosis (**117-118-119**). Other M proteins act as antiphagocytic factors by binding to fibrinogen (**120**), such as M1 protein that binds fibrinogen via the B-repeats (**121**, 1.7.1.1). This interaction diminishes covalent binding of the C3 convertase of the complement system on the streptococcal surface, resulting in reduced cleavage of C3 into C3a and C3b. C3a is an anaphylatoxin that attracts and stimulates immune cells (**122**). The C3b opsonizes *S. pyogenes* for phagocytes that harbour the C3 receptor. Hence, interfering with C3b deposition by binding fibrinogen reduces phagocytosis (**103**).

In addition, M protein, released from the streptococcal surface, forms complexes with fibrinogen. These aggregates activate neutrophils by binding to β_2-integrin of these cells. As a result, neutrophils release heparin binding protein, an inflammatory mediator that induces vascular leakage. In mice, injection of M protein or subcutaneous infection with *S. pyogenes* causes severe pulmonary damage characterized by leakage of plasma and blood cells, which resembles the pathology of STSS (**123**). Moreover, M protein/fibrinogen complexes were identified in tissue biopsies from a patient with necrotizing fasciitis and STSS, further underlining the pathogenic significance of such complexes in severe streptococcal infections (**124**).

Invasive *S. pyogenes* infections cause severe systemic inflammatory reactions and complications such as sepsis (**28-124**). Important triggers of inflammation include streptococcal superantigens that evoke a strong cytokine production by activating T cells (**124**), streptolysin O that stimulates human neutrophils (**125**) and M protein (**126**). In

addition, M protein stimulates infiltrating monocytes/macrophages, inducing secretion of IL-6, IL-1β and TNFα (**126**). Highly purified M1 protein induced T-cell proliferation and release of TNFβ and INFγ, triggering inflammatory processes (**127**). Taken together, the M protein contributes to local and systemic inflammatory reactions during streptococcal infections.

In addition to their role in acute infection, M proteins are suspected triggers of ARF as described in paragraph 1.7.2.

1.6.2 M-like proteins

Approximately half of all *S. pyogenes* isolates have up to two genes that code for M-like proteins, the *emm*, the *mrp* and the *enn* genes (**128-129-130-131**). These genes are adjacent to the *emm* gene in the Mga region of the streptococcal chromosome and are under the control of the positive regulator Mga (multi gene activator) (**132-133-134**) (Fig. 3). The respective M-like proteins are referred to as M-related protein (Mrp) and Enn, which are immunologically dissimilar (**135**).

Like some of the M proteins, Mrp is a streptococcal IgG-receptor that binds to the Fc region of the immunoglobulin. For this reason it is also named fcR (Fc receptor). As it binds to fibrinogen, like M proteins, it is considered as an antiphagocytic factor (**136**). Moreover, the co-operation of M proteins and Mrp contributes to phagocytosis resistance by decreasing streptococcal binding to granulocytes (**136**).

The majority of *enn* genes encode proteins that bind IgA (**137**). *In vitro* the *enn* gene is transcribed at a very low level when compared to the *emm* and *mrp* genes, or not at all (**137-138-139**). However, the expression of *emm*, *mrp* and *enn* gene products can vary depending on infection sites and may be part of the bacterial strategy to evade the immune system and colonize the host (**137**).

Studies on *emm*, *enn* and *mrp* deletion mutants showed an increased sensitivity to phagocytosis as compared to the wild type strain (**140**). These data suggested a functional redundance among M and M-like proteins that provides resistance against phagocytosis (**141**). An extensive knowledge about the role of Mrp and Enn in pathogenesis is missing.

A study on streptococci of *emm*-type 18 identified a surface protein called "streptococcal protective antigen" (Spa18), (**142**). This *emm*-type was responsible for an endemic outbreak of *S. pyogenes* infections that led to ARF (**143**). An analogous investigation on streptococci of *emm*-type 36 identified the allelic variant Spa36 (**143**). Spa18 and 36 share characteristics with M-like proteins, but the *spa* gene that codes for these proteins is not located in the Mga region. Like the M proteins and the M-like proteins Mrp and Enn, Spa proteins are hypervariable in their N-terminal sequence and the C-terminus is conserved. Spa proteins Spa18 and 36 are up to 95% identical to the M-like protein of *S. equi* subsp. *zooepidemicus* suggesting horizontal transfer of the gene to *S. pyogenes* (**65**). Secondary structure prediction using Paircoils2 identified a long central coiled-coil region in Spa (**65**). Spa18 is located on the streptococcal surface and together with the M protein contributes to the virulence of streptococci by acting against phagocytosis (**141-143-144-145**). In conclusion, M-like proteins Mrp, Enn and Spa represent a class of streptococcal proteins that share structural and functional homologies with the M protein. Similar to M proteins, they may have a role in ARF, which remains to be investigated.

1.7 Streptococcal autoimmune sequelae

1.7.1 Poststreptococcal glomerulonephritis

Poststreptococcal glomerulonephritis (PSGN) is an immune sequela that appears with a delay of 1 to 3 weeks after infection by *S. pyogenes* or *S. equi* subsp. *zooepidemicus.* This disease occurs primarily in children, young adults and individuals over the age of 40 years (**146**).

PSGN is associated with pyoderma in tropical climates throughout the year and in temperate climates during the summer season (**147**). Alternatively, during cold seasons PSGN is associated with throat infection in regions with temperate climate (**146**). Crowding and poor hygiene promote PSGN outbreaks (**148-149-150**). Moreover, in *S. pyogenes* the nephritogenic potential seems to be linked to certain *emm*-types, which suggests a role of the M protein or other type specific factors in triggering PSGN (**150**). Early studies suggested shared epitopes between streptococci and glomeruli as trigger of the disease (**152-153-154**). However, an extensive review of the experiments reporting cross-reactivity between mammalian tissues and streptococci concluded that most of the preparations used in these studies were likely to contain streptococcal IgG receptors or tissue IgG Fc receptors (**155**), which invalidated the hypothesis of molecular mimicry.

Histopathological examination of afflicted glomeruli revealed characteristic structures called sub-epithelial ‘‘humps’’ on the luminal site of the glomerular capillary basement membrane. Streptococcal antigens, complement factors and immunoglobulin are localized in these humps (**156-157**). This supports the notion of immune-mediated mechanisms in the pathogenesis of PSGN (**158**). In fact, circulating immune complexes containing streptococcal antigens were found in 58% of patients with PSGN (**158-159**). Interestingly intact *S. pyogenes* was never found in glomeruli of patients with PSGN (**160**).

Of the variety of potential nephritogenic factors that have been under investigation, so far there is none that can be considered as the sole nephritogenic factor. Despite the fact that many of the suspected nephritogens accumulate in the kidney of laboratory animals, leading to glomerular injury, none of these factors alone was capable to evoke a disorder that resembled PSGN sufficiently. The repertoire of host pathogen interactions that leads to PSGN is not yet uncovered and the molecular pathogenesis of PSGN is still widely elusive.

1.7.2 Acute rheumatic fever and rheumatic heart disease

1.7.2.1 Epidemiology of ARF and RHD

Acute rheumatic fever (ARF) is an immune sequela of insufficiently treated streptococcal sore throat that can lead to rheumatic heart disease (RHD). Most of the patients with ARF are children, adolescents or young adults. More than 500.000 new cases of ARF occur every year (**26**). RHD has a prevalence of at least 15 million cases worldwide, with 282 000 new cases each year (**26**). ARF and RHD are responsible for about 200.000 deaths every year. Some geographic regions e.g. Sub-Saharan Africa, South central Asia and some communities of Indigenous Australians are particularly affected by a high prevalence of ARF/RHD (**28-50-159**). 95% of the estimated deaths caused by ARF or RHD occur in these regions (**26**), where RHD remains the major acquired cardiologic disease in the young and middle age population (**26**). Prevalence of ARF and RHD is high among people that live under low socioeconomic conditions and have poor access to medical care. Moreover, certain humans appear genetically predisposed to ARF as shown by association with certain HLA types (**28**). However, the value of the HLA type as a prognostic marker for ARF is not clear (**161**).

1.7.2.2 Pathology of ARF

ARF occurs with a latency period of 1–4 weeks (**162**). The diagnostic criteria for ARF are known as the Jones criteria, which comprise

Major Jones criteria:

- Carditis (murmur, cardiomegaly, pericarditis, and congestive heart failure)
- Erythema marginatum
- Migratory large joint polyarthritis
- Sydenham chorea
- Subcutaneous nodules

Minor Jones criteria:

- A previous history of rheumatic fever
- Arthralgia
- Fever

Positive diagnosis of rheumatic fever is made when at least two major or one major and two minor Jones criteria are met (**163**). This diagnosis is strongly supported by evidence of a preceding streptococcal infection, e.g. elevated antibody titers against streptococcal antigens like streptolysin O. About a third of the ARF patients develop RHD, leading to irreversible damage of the heart valves (**159**). These patients require surgical intervention and often substitution of the heart valves.

Histopathologically, ARF is characterized by inflammatory changes in the extracellular matrix or subendothelial and perivascular tissue (**164-165**). Granuloma formation comprising perivascular Aschoff nodules has been described as the most-characteristic finding in rheumatic carditis (**166-167**). The depth of perivascular inflammation is limited and interstitium beyond this area appears normal. In almost all patients with ARF the pericardium is affected. The myocardium is often spared (**164**) and no residual damage in myocardium or

pericardium is observed after the acute episode of rheumatic fever is resolved (**164**). In contrast, the valvular tissue sustains permanent damage after active carditis (**164**). The mitral valve is most frequently affected (**167**), followed by the aortic valve that is frequently inflamed (**164**).

In conclusion, ARF pathogenesis suggests that the primary site of damage is the subendothelial and perivascular extracellular matrix (**166**).

1.7.2.3 Pathogenesis of ARF

Elevated autoantibody titers in the sera of ARF/RHD patients (**101-113-162**) and deposition of immunoglobulin and complement in the affected hearts (**168**) indicate that humoral immunity is involved in ARF pathogenesis. The early processes in rheumatic carditis involve cellular immune responses as indicated by the high number of T-lymphocytes and macrophages that occur in the cellular infiltrates observed in myocardial lesions of affected hearts (**169-170-171**). Moreover, the inflammation that takes place in those patients is characterized by increased numbers of mature circulating T helper cells (**170**), elevated levels of cytokines such as IL-1 and IL-2, IFN-γ, TNF-α and higher amounts of receptors for these cytokines (**170**). Autoimmune responses against several host proteins such as cardiac myosin, heart valve glycoproteins, antigens of the sarcolemma membrane and collagen occur in ARF or RHD patients (**101-170-172-173**). These observations suggest that ARF is caused by detrimental immune responses and inflammation that are not sufficiently suppressed after infections caused by *S. pyogenes*.

1.7.2.4 Streptococcal factors in ARF

The molecular and immunological processes that could drive the pathogenesis of ARF and subsequent RHD have been partially discovered. (**101-170-174**). Streptococcal factors such

as streptococcal superantigens, group A carbohydrate, hyaluronic acid capsule and M proteins have been suspected to trigger the autoimmunity that is seen in ARF. The role of these bacterial components in the induction of ARF is described below.

1.7.2.4.1 Superantigens in ARF

During acute infection, streptococcal superantigens contribute to virulence by activation of T cells and induction of strong inflammatory responses (**175-176-177**). Superantigens bridge the MHC class II of the antigen presenting cells with the T-cell receptor, which activates the T-cells that carry a matching T-cells receptor V-beta region. These T-cell populations will release pro-inflammatory cytokines. Furthermore, interaction of coreceptors CD28 or LFA-1 on T-cells with the ligands B7 or ICAM-1 on antigen presenting cells, stimulate a polyclonal expansion of the activated T-cells. In the absence of these co-stimulating interactions, the activated T-cells fall into anergy or apoptosis (**146**), resulting in the selection for a certain T cell population. It has been speculated that if the mitogenic activity induced by superantigens is combined with a break of tolerance, the patient may produce auto-reactive cells (**176-178**). In this case, superantigens may play a role in the autoimmune processes seen in ARF pathogenesis that remains to be further investigated.

1.7.2.4.2 Group A carbohydrate in ARF

Crossreactivity of autoantibodies from ARF patients with streptococcal cell wall carbohydrates, the cell membrane or the M protein was demonstrated in different studies (**179-180**). Therefore, monoclonal antibodies (mAbs) have been used to characterize the crossreactivity between streptococci and the host proteins, which contributed to the understanding of the pathogenesis of rheumatic heart disease. Monoclonal antibodies against N-acetyl-beta-d-glucosamine (GlcNAc), the immunodominant epitope of group A carbohydrate, have been detected in patients with rheumatic valvular disease suggesting their

role in the pathogenesis (**181**). Notably, GlcNAc is ubiquitous in the host's extracellular matrix and shares epitopes with the streptococcal group A carbohydrate that contains the same monosaccharide unit (**171**). Thus, autoimmune responses against GlcNAc of the group A carbohydrate may be causative for the early, less specific symptoms in ARF (**159**).

1.7.2.4.3 Role of hyaluronic acid in ARF

Hyaluronic acid is produced by the bacteria as well as it is abundant in the host's extracellular matrix. *S. pyogenes* M18, an isolate that is highly encapsulated with hyaluronic acid, caused two outbreaks of ARF in Utah, USA (**182**). Hence, the *S. pyogenes* capsule was suspected to cause autoimmunity. More recent work suggests that the hyaluronic acid capsule binds and aggregates collagen on the streptococcal surface, which could evoke the production of collagen autoantibodies similar to what has been observed for M protein (1.7.2) (**113**). However, isolated hyaluronic acid alone seems not sufficient to trigger disease, since it is widely used in cosmetic medicine with a low risk for immune complications.

1.7.2.5 Role of M protein in ARF

1.7.2.5.1 M protein and molecular mimicry

M proteins form dimers via long α-helical coiled-coil regions. Coiled-coil structures are also found in several host proteins such as α-keratin, human heart myosin and laminin. This structural similarity between host molecules and the M protein is the cause of immunological crossreactivity known as molecular mimicry (**183**). Molecular mimicry is considered as one of the triggers of the detrimental immune responses in ARF (**174-184**). The blood of ARF/RHD patients contains higher concentrations of these cross-reactive antibodies as compared to healthy individuals (**185**). Notably, the autoantibodies from ARF patients differ in reactivity from anti-cardiac antibodies seen in heart failure or postcardiotomy patients, which are a sequel of cardiac damage (**186**). This suggests that the autoimmunity in ARF is

the cause for the cardiologic symptoms. Monoclonal antibodies derived from ARF/RHD patients were poly-reactive to M protein, human coiled-coil proteins and GlcNAc (**28**). Human myosin cross-reactive antibodies recognize epitopes in M proteins M5, 6 and 19 (**187-188**) and appear to contribute to the pathogenesis of carditis in ARF (**28**). Therefore, molecular mimicry between streptococcal M protein and host tissue is considered as one of the initiating mechanisms of this disease. In addition, T-cells isolated from severe cases of RHD cross-recognized M5 peptides, heart tissue proteins and myosin peptides. This stimulates the T cells to produce high amounts of inflammatory cytokines and low levels of IL-4, suggesting that mimicry between streptococcal antigens and heart-tissue proteins, combined with high inflammatory cytokine and low IL-4 production, leads to the development of autoimmune reactions and cardiac tissue damage in RHD patients (**189**). Taken together these experimental data indicate molecular mimicry as one cause for autoimmunity that leads to ARF (**174-184**).

1.7.2.5.2 M protein, PARF and collagen binding as a trigger of ARF

An alternative mechanism of triggering autoimmune responses that are observed in ARF involves the M protein's ability to induce autoimmune response against collagen IV (**101-102 113**). The M3 protein (**190**) that caused ARF and ARF-like symptoms in vaccinees (**71)**, binds collagen to the streptococcal surface and causes its aggregation (**113**). This interaction, which occurs with collagen of types I–IV, has been observed in different types of M protein including M proteins of SDSE and it depends on an octapeptide motif located in the N-terminal hypervariable part of the M protein (Fig. 6) (**59-101-102**). This motif contributes to acute infections by facilitating streptococcal colonization of the extracellular matrix (**113**). M3 protein and other collagen-binding M proteins that carry the PARF motif trigger production of anti-collagen IV antibodies. Therefore, it has been suggested that induction of collagen autoimmunity requires the binding of M protein to collagen. Detrimental

autoimmunity against collagen may initiate the inflammation seen in ARF. Collagen IV is a major component of basement membranes and is localized beneath the endothelium that covers the heart valves. Thus, it could be one of the target of the autoimmune responses that lead to valvular damage. This hypothesis is supported by the fact that anti-collagen IV autoantibodies were also found in increased levels in ARF and RHD patients as compared to healthy individuals (**101-113**). For this reason, the collagen binding motif was named PARF, as an abbreviation of peptide associated with rheumatic fever (**101**). The increase in collagen IV autoimmunity occurs early during acute pharyngitis before the onset of ARF, which indicates that the anti-collagen IV immune response is causative for the sequelae rather than a consequence of the tissue destruction seen in ARF and RHD (**113**).

Different from the myosin-directed autoimmunity, collagen IV autoantibodies produced in response to PARF positive M proteins do not cross-react with the M protein, suggesting that the immune reactions against collagen in ARF/RHD are not caused by molecular mimicry. The binding of M protein to collagen may alter the presentation of the autoantigen, leading to a break of tolerance (**101-113**). The detailed pathomechanisms, however, remains elusive. The PARF-dependent collagen autoimmunity may be a ‘‘conformeropathy’’ analogous to the Goodpasture’s syndrome. In Goodpasture’s syndrome, a conformational change in the non-collagenous domain of collagen IV leads to exposure of epitopes that evokes the production autoantibodies against the glomerular basement membrane. This causes a collagen IV autoimmune disease that affects the lungs and kidneys (**191-192**).

The *S. pyogenes* of *emm*-type 3, which harbour PARF, represent 7% of the *S. pyogenes* isolates from human infections world-wide (**193-194**) which underlines their clinical relevance.

Although the PARF motif binds to a variety of collagens, e.g. collagens I and IV, not all the PARF-containing M proteins bind collagen with high affinity (**59**). This creates an opportunity to investigate the role of PARF and its ability to bind collagen in triggering the collagen autoimmune response. This can for the first time shed light on the role of a protein interaction in induction of autoimmunity.

S. pyogenes possesses several collagen-binding proteins that do not belong to the M- or M-like proteins. The protein Cpa (collagen-binding protein of group A streptococci) that bound to collagen type I, was identified in about 30% of 68 different *S. pyogenes* serotypes (**195**). In addition, the streptococcal leucine rich (Slr) protein, an abundant membrane bound lipoprotein, is co-expressed on the bacterial surface with M1 and was shown to be involved in the interaction with collagen type I (**196**). However, the role collagen binding proteins like Cpa and Slr in triggering collagen autoimmunity remains unexplored.

1.7.3 Role of *S. dysgalactiae* subsp. *equisimilis* in ARF

SDSE is a major human pathogen of the GCGS (**197-198**). Moreover, there are convincing indications that infection with SDSE can cause ARF (1.4 SDSE), which is challenging the previous picture of the etiology of ARF.

The vast majority of SDSE strains that colonize or infect humans have M proteins like *S. pyogenes*. This explains why pharyngeal SDSE isolates from a community of Indigenous Australians evoked anti-myosin responses in mice. Most likely, this is reflecting a rheumatogenic potential of SDSE strains that is responsible for the high prevalence of ARF in this community (**44**). More than 10% of clinical SDSE isolates collected in Southern India bore collagen-binding M proteins. This region is also known to suffer from a high prevalence of ARF (**59**) as well as high rates of SDSE throat carriage and pharyngitis (**199**). Taken together, SDSE carry the factors that trigger ARF-associated autoimmunity. Although the

relevance of SDSE in ARF requires further examinations, these bacteria must not be neglected in the medical practice and in research as potential cause of ARF.

1.7.4 Outlook

Discovering the factors and the molecular mechanisms by which streptococci trigger autoimmune sequelae will help to combat this significant threat to human health. The identification of streptococci with a nephritogenic or rheumatogenic potential could facilitate identification of patients who are at high risk of contracting these immune diseases and who benefit from an intensified preventive treatment. Moreover, increasing knowledge on rheumatogenic streptococci and their rheumatogenic factors may have a substantial impact on the design of vaccines that aim at preventing the infections and their autoimmune sequelae. Future vaccines may have to protect against a broader range of streptococci that includes SDSE.

1.8 Vaccines against *S. pyogenes*

The substantial impact of *S. pyogenes* on human health has motivated more than half a century of search for a vaccine against this pathogen. Although several *S. pyogenes* vaccine formulations elicit protective immunity in animal models and have shown promising results in early-stage human trials, a safe and effective vaccine against *S. pyogenes* for use in humans is not yet available (**28-200**).

In order to develop a vaccine that covers a broad variety, if not all *S. pyogenes* bacteria, several antigens have been investigated as potential vaccine candidates. However, several obstacles were encountered during the development of such a vaccine.

As described in paragraph 1.5, *emm*-typing studies shed light on the diversity and the regional and temporal variation of circulating *S. pyogenes* genotypes (**78**). The corresponding high antigenic diversity of *S. pyogenes* reduces the number of conserved protective antigens and is one of the obstacles in vaccine development. Several approaches have been adopted to overcome this problem. These approaches include the design of multivalent vaccines that incorporate several streptococcal antigens or the design of vaccines that target the circulating strains of a specific geographic region (**200**).

The development of a vaccine is further complicated by serotype replacement. For instance, the replacement of type *emm*1 by *emm*6 was observed under the pressure of protective immunity against *emm*1 within a short period of time (**78**). This indicates that a vaccine that targets only a few serotypes could be effective just for a short period of time, after which new serotypes could replace the types that are covered by the vaccine.

As natural streptococcal infections or preparations of streptococcal antigens cause ARF, streptococcal vaccines are associated with a risk of induction of one of the diseases that they should prevent (27, 1.8.2). This problem goes alongside with the lack of animal models that allow to foresee the immune side effects of the vaccine (**65**).

While facing these obstacles, vaccinology has followed two strategies in the development of *S. pyogenes* vaccines: M-protein-based vaccines and non-M protein based vaccines.

1.8.1 Non-M protein based vaccines

Because the M protein is suspected to trigger autoimmune side effects, research moved towards the design of non-M protein based vaccines (**71,** 1.6.1). The aim of this research is to find a safe vaccine formulation that is effective against a broad variety if not all *S. pyogenes*

bacteria. Some examples of these *S. pyogenes* vaccine candidates are described in this chapter.

1.8.1.1 Group A carbohydrate

As all known isolates of *S. pyogenes* carry the Lancefield group A carbohydrate (GAC) on their surface, this macromolecule was one of the first isolated antigens that were studied for use in a streptococcal vaccine. Anti-GAC antibodies opsonized *S. pyogenes* and protected immunized mice against challenge with *S. pyogenes* of different serotypes (**201-202**). However, immunological crossreactivity between anti-GAC antibodies with host proteins of the heart valve and the cytoskeleton, such as actin, keratin, myosin and vimentin (**174**) prevent the use of GAC as a *S. pyogenes* vaccine constituent. (**200-203**).

1.8.1.2 Streptococcal protective antigen

A more recent vaccine candidate is the M-like protein streptococcal protective antigen (Spa) (1.6.2). Spa elicited bactericidal antibody responses in animals (**141-143**). Spa-specific antiserum reacted with Spa or Spa-like proteins on streptococci of 25 out of 70 different *emm*-types, suggesting that the recognized epitopes are conserved among a variety of *S. pyogenes* isolates (**143**). An N-terminal fragment of Spa is a component of the 26-valent M protein based vaccine named StrepAvax^{TM} (ID Biomedical Corporation, Canada), which has completed a phase II clinical trial and of the successor 30-valent vaccine, which is currently in a preclinical phase (1.8.2).

1.8.1.3 Streptococcal cysteine proteinase A

Streptococcal cysteine proteinase A (SCPA) is a sortase-anchored surface protein of *S. pyogenes*. This enzyme is also referred to as C5a peptidase as it cleaves the anaphylatoxin

C5a of the complement cascade, acting as an anti-phagocytic factor (**204-205**). Thereby, it promotes the virulence of streptococci (**206**). Homologous C5ase enzymes are also produced by streptococci of groups B, C and G, indicating that a SCPA vaccine could cross-protect against infections with these β-hemolytic streptococci as well (**207**). An injectable formulation of SCPA facilitated clearance of intranasally inoculated *S. pyogenes* in mice (**208**). A SCPA-based vaccine could cover the great majority of *S. pyogenes* strains as this antigen is carried by most, if not all isolates (**209**). Moreover, as SCPA has not yet been implicated in the induction of autoimmunity, it remains a promising vaccine candidate (**210**).

1.8.1.4 Streptococcal superantigens

A streptococcal vaccine that uses inactivated streptococcal superantigens SpeA or SpeC, evoked neutralizing antibody responses against these toxins. These responses protected rabbits against the toxic effect of the respective isolated superantigen (**211-212**). During acute streptococcal infection, immune responses against these superantigens may protect also humans from the toxic effects of the secreted superantigens, but the immune response may be non-bactericidal. In fact, toxin specific IgG responses may not be sufficient to opsonize the pathogen and prevent the adaptive immune responses against *S. pyogenes* that lead to ARF. Moreover, although it has been speculated about a role of superantigens in the induction of ARF (**213**), there are no experimental indications that they trigger autoimmunity (**214-215**) and their role in ARF pathogenesis is not sufficiently investigated.

1.8.2 M protein based vaccines

M protein based vaccines strive to exploit immune responses against this streptococcal cell wall-anchored protein. This is motivated by the protectivity of these responses in animal models of acute infection, as well as evidence for their protectivity in humans (**216**).

1.8.2.1 Potential risks associated with M protein vaccines

Immunizing healthy human volunteers with preparations of the M protein decreased their colonization with life *S. pyogenes* after pharyngeal challenge (**217**). However, M3 protein of *S. pyogenes*, caused ARF or ARF-like symptoms in three out of 21 vaccinated children in early trials in the late 1960s (**71**), significantly more than in the untreated population. The potential of this antigen to cause autoimmune side effects in children underscores the role of full-length M proteins as triggers of autoimmunity.

Several studies identified similarities in the coiled-coil structure of both M protein and myosin and suggested this molecular mimicry as the trigger of autoimmunity (**169-189**). Thus, a risk factor in M protein based vaccines. The myosin crossreactive sequence (Gln–Lys–Ser–Lys–Gln) was identified in M5 and M6 proteins and shown to be located near the pepsin cleavage site (**188**), after amino acid 228, in the B5 repeat of the M protein (**79**). Moreover, other myosin crossreactive sites were identified in M5 (**217)** and M19 (**187**). Using overlapping synthetic peptides of M5 protein, myosin crossreactive B-cell epitopes were identified in peptides from the A, B and C repeat regions of M5 protein (**218**). Several approaches have been taken to exclude such parts from M protein based vaccine that could be the cause for immunological side effects. On the way to develop a safe vaccine it will be important to identify and exclude other rheumatogenic parts of the M protein. One of the possible rheumatogenic parts of the M protein is the PARF motif (**101**, 1.6.1).

1.8.2.2 N-terminal M protein based vaccines

One of the most promising vaccine candidates is based on the N-terminal hypervariable region of the M protein, a 26-valent vaccine named StrepAvax (**219**). This vaccine is composed of four different fusion proteins that contain six to seven N-terminal M protein fragments (50 amino acids) in a concatenated arrangement. The vaccine utilizes epitopes of M

proteins that evoke the production of bactericidal antibodies and that are less likely to cross-react with human tissues (**218**). StrepAvax was designed to cover up to 26 different *emm*-types, but it cross protected against other *emm*-types as well (**220**). The vaccine covers the predominant *emm*-types that caused invasive infections or ARF, in the USA and in Europe. In a human phase I clinical trial, StrepAvax evoked *ex vivo* bactericidal responses against all of the *emm*-types that were included in the vaccine (**220**). The vaccine did not produce ARF or ARF-like symptoms in 30 adult volunteers (**220**) despite the presence of PARF in the M3 specific part of one of the vaccine antigens.

Because ARF is mainly affecting children and adolescents, a vaccine against *S. pyogenes* needs to be safe for use in these age groups, which might react different from adults. Notably, the ARF-like symptoms that followed upon vaccination of 21 children with M3 protein, occurred only in children that were reinfected by streptococci (**71**). Thus, reinfection may be crucial for occurrence of the autoimmune side effects. As streptococcal pharyngitis is more frequent in children, they may have a higher risk to develop such side effects upon vaccination with StrepAvax (**65**). Data about the safety of StrepAvax from the phase II clinical trial on children are not yet available.

With the aim of increasing the *emm*-type coverage, a 30-valent vaccine was designed and subjected to preclinical tests. This vaccine evoked bactericidal antibodies against 28 of the 30 targeted *emm*-types in rabbit. Moreover, bactericidal cross-reactivity with 24 out of 40 non-vaccine *emm*-types was observed, indicating that the actual *emm*-type coverage of this multivalent M protein vaccine is higher than its valence (**221**). Still, StrepAvax and the 30-valent vaccine may cover the predominant *emm*-types that caused infections in the USA and in Europe but may not be sufficient to combat ARF/RHD in regions with a high prevalence of these diseases due to low *emm*-type coverage rates (**222**, 1.6.1). Furthermore, serotype

replacement, observed in different regions of the world, may be a further threat to the success of these vaccine candidates.

In addition, the multivalent vaccines include in their formula the N-terminal part of the M3 protein, which contains the PARF motif. As this motif could to be involved in triggering rheumatogenic processes, the multivalent vaccines in their present form may be not safe.

1.8.2.3 C-terminal M protein based vaccines

Certain N-terminal parts of the M protein are suspected to trigger autoimmune reactions due to cross reactive epitopes, as explained in the paragraph 1.7.2 (Role of M protein in ARF). Therefore, research moved towards vaccines that use a more C-terminal part of the M protein that is conserved between different M-types, named conserved C-repeat region (CCR). Hence, the CCR could yield a better *emm*-type coverage as compared to peptides of the hypervariable part of M protein such as the peptides used in StrepAvax. Early studies showed that CCR derivative peptides were protective in mice that were challenged with *S. pyogenes* of different M types (**223-224**).

One of the vaccines that exploits the CCR of the M protein is the StreptInCor vaccine (Fig. 8). This vaccine uses a peptide of 55 amino acids that combines a protective T-cell and B-cell epitope (**225**). StreptInCor induced high titers of specific antibodies as well as specific T-cell immune responses in mice. Immunized mice exhibited a high survival rate one month after challenge with *S. pyogenes*. Moreover, no cross-reaction to cardiac myosin was detected. Taken together, these data underline that StreptInCor could be an effective and safe vaccine for prevention of *S. pyogenes* infections (**226-227-228**).

Figure 8: **Alignement of four CCR vaccines candidates,** (names are given in bold on the left). The minimal common T-cell epitope is marked by a box (**227-229-230-231**).

An alternative to StreptInCor is the 20 amino acid peptide within the CCR, named p145 (**229**). Because a high proportion of people living in areas with high prevalence rates of ARF and RHD had increased titers of antibodies against p145 in their sera (**232**), this antigen has been investigated for use in a vaccine that could be protective against ARF and RHD.

A derivative of p145 is the peptide J14. J14 is a part of p145 that contains a minimal T-cell epitope and carries heterologous flanking regions that preserve the α-helical structure of the internal M protein epitope, which is referred to as J14i (**231**). Vaccination with J14 protected mice that were challenged with *S. pyogenes* and led to production of bactericidal antibodies (**233**). J14i variants were identified in 77 different *emm*-types (**234**). The antigen SV1 that was composed of seven J14i modules of five J14i variants evoked production of bactericidal antibodies in mice (**234**). Moreover these antibodies were bactericidal against *S. pyogenes* serotypes that were not included in the vaccine *in vitro* (**234**). As for the multivalent N-terminal M protein vaccines, despite promising preclinical testing, the available data do not yet fully exclude potential autoimmune side effects. However, due to the absence of known cross-reactive epitopes and lack of a PARF motif, such an antigen may be a favourable starting point for the development of a safe and effective vaccine against *S. pyogenes*. A further strength of the CCR-based vaccine as compared with StrepAvax is the use of less and similar antigen modules to cover the variety of *S. pyogenes* serotypes. A general CCR vaccine may be easier to design than an N-terminal M protein vaccine and bear a lower risk of

including rheumatogenic factors in the multivalent vaccine. Moreover, the risk of losing efficacy due to serotype replacement may be lower for CCR vaccines.

1.8.3 Perspectives in development of a *S. pyogenes* vaccine

Several vaccine candidates have been identified in the past and have shown promising results in preclinical models. Still, a vaccine that is protective against the acute infections and sequelae caused by *S. pyogenes* is not yet available and it cannot yet be foreseen when a streptococcal vaccine will be licensed. Major difficulties in the development of a vaccine are determined by the nature of streptococcal antigens used as vaccine candidates, as described in this chapter (1.8). However, the growing knowledge on these antigens and comprehensive identification of their noxious parts, which cause side effects, may contribute to guide the development of safe and effective vaccines against ARF. In addition to this, an important help to vaccine development against streptococcus can be given by reverse vaccinology (**235**). Reverse vaccinology is a new approach for vaccine development that aims to identify the complete repertoire of antigens that a pathogen expresses on its surface by screening of the whole genome sequence (**236**). This new approach can be of great help to find new protective streptococococcal antigens and contribute to the development of an effective streptococcal vaccine.

1.9 Aim of this study

Previous studies indicated that the PARF motif is a possible player in triggering autoimmunity against collagen IV, an autoimmune response that occurs in ARF patients. It had not been studied if other motifs than PARF or other functions than collagen binding of the M protein are involved in triggering autoimmunity against collagen IV. Therefore, the aim of this study was to investigate the specificity of PARF as a trigger of collagen autoimmunity and the role of the collagen binding activity of PARF in this process. Different M proteins with and without PARF motif and differing in affinity for collagen IV should be produced recombinantly for this purpose and tested in a mouse model for induction of autoimmunity. Moreover, to test isolated PARF motifs, a synthetic peptide (peptide 17) that was optimized for collagen binding and recombinant protein that carried all known naturally occurring PARF motifs (PARF tandem) should be examined. As PARF tandem could be a useful vaccine antigen for immunization against PARF positive strains, one aim was to test its ability to evoke the production of opsonizing antibodies. PARF is present in about 7% of *S. pyogenes* isolates from human infections. In addition, other collagen binding factors may exist in these bacteria that trigger collagen IV autoimmunity. The role of Spa18, an M-like protein, in collagen binding and as a trigger of autoimmunity, should therefore be tested. Spa18 is present im *emm*18 strains of *S. pyogenes*, which have caused several outbreaks of ARF.

The results should deliver new insights into the elicitation of autoimmunity by streptococcal proteins.

CHAPTER 2

Materials and Methods

2.1 Buffers and solutions

- **Bicarbonate coating buffer:** 0.1 M $NaHCO_3$ in DI water adjusted to pH 9.0.
- **Luria-Bertani medium (LB):** tryptone 10 g, yeast extract 5 g, NaCl 10 g in 1 L DI water.
- **Luria-Bertani agar (LA):** LB plus 1.5% (wt/vol) agar.
- **Phosphate buffered saline (PBS)**: 8 g of NaCl, 0.2 g of KCl, 1.44 g of Na_2HPO_4, 0.24 g of KH_2PO_4, in 1 L DI water, adjusted to pH 7.4.
- **PBST:** PBS containing 0.05% (vol/vol) of Tween 20.
- **Super optimal broth with catabolite repression (S.O.C.):** tryptone 20 g, yeast extract 5 g, NaCl 0.5 g, KCl 0.18 g, $MgCl_2$ 0.95 g, glucose 3.6 g, in 1 L Milli-Q water, adjusted to pH 7.0.
- **Tryptic soy broth (TSB):** tryptone 17.0 g, soytone 3.0 g, glucose 2.5 g, NaCl 5.0 g, K_2HPO_4 2.5 g, in 1L DI water, adjusted to pH 7.3.
- **TSS storage and transformation solution (TSS):** LB with 10% (wt/vol) polyethylene glycol (PEG molecular weight 8000), 5% (vol/vol) dimethyl sulfoxide (DMSO) and 20-50 mM Mg2+ (MgSO4), adjusted to pH 6.5.
- **Todd-Hewitt broth supplemented with 0.5% yeast extract (THY):** heart infusion 3.1 g, neopeptone 20.0 g, yeast extract 5.0 g, dextrose 2.0 g, NaCl 2.0 g, Na_2HPO_4 0.4 g, Na_2CO_3 2.5 g, in 1L DI water, adjusted to pH 7.8.

2.2 Ligation and Endonuclease digestion

DNA ligations were performed using T4 DNA Ligase (Thermo Scientific) and following the manufacturer's protocol.

Endonuclease digestion of plasmids and PCR amplicons (25 µL) were performed using 1 µL FastDigest™ enzyme *BamHI*, *SalI* or *EcoRV* and *5* µL FastDigest™ Buffer (Thermo Scientific), in a volume of 50 µL for 1 h at 37°C. 1 µL of enzyme solution corresponds to 1 FDU, which cleaves 1 µg of lambda DNA/Bsp120I fragments in 5 min at 37°C in FastDigest™ Buffer.

2.3 Isolation and purification of the DNA

PCR amplicons, plasmid and endonuclease digested DNA were isolated and purified with the NucleoSpin® Gel and PCR Clean-up Kit (Macheray-Nagel) following the manufacturer's instructions.

2.4 Preparation of competent cells

Competent cells were prepared as described by Chung (**237**) with some modifications. Using a toothpick, *E. coli* cells were plated on LA containing antibiotics (100 µg/mL ampicillin for *E. coli* DH5-α or 100 µg/mL ampicillin and 25 µg/mL kanamicin for *E. coli* M15). Bacteria were incubated overnight at 37°C. On the following day, bacteria were inoculated in 10 mL of LB medium with antibiotics (see above LB cultures) and cultivated overnight at 37°C in a shaking incubator at 120 rpm. On the third day, 20 mL of culture was added to 200 mL of prewarmed TSS solution (2.1) until the bacteria reached the early exponential phase at an optical density (O.D.) of 0.3-0.4 at 600 nm wavelength. The bacteria were further cooled in ice for 10 min, centrifugated at 1000 x g for 10 min at 4°C and subsequently resuspended in

ice-cold TSS solution to one-tenth of the original volume. Aliquotes of 100 µL of bacteria in TSS were prepared in 1.5 mL Eppendorf Tubes®. The aliquots were quickly frozen in liquid nitrogen and stored at -80°C.

2.5 Transformation of competent *E. coli* bacteria

Aliquots of 100 µL of competent cells (*E. coli* M15 or DH5-α) were taken from -80°C and left for 5 min in ice. 1 µL of plasmid DNA, or 7 µL of ligation product were added and the bacteria were incubated on ice for further 30 min. After a heat shock step at 42ºC for 45 s in a water bath, the *E. coli* were incubated on ice for 10 min. Then, 300 µL of pre-warmed LB or S.O.C. medium without antibiotics was added and the bacteria were incubated at 37ºC for 1h in an incubator-shaker at 300 rpm. Finally, 200 µL of culture was plated onto pre-warmed LA with antibiotics (ampicillin 100 µg/mL) for DH5-α (ampicillin 100 µg/mL and kanamicin 25 µg/mL) for *E. coli* M15 and incubated at 37ºC overnight. Positive clones were analyzed by colony PCR using the following conditions: 1 unit of Taq DNA Polymerase (Qiagen), 10 pmol of primers 18 and 19 (Table 1), 1.25 mM $MgCl_2$ and 0.04 mM of each dNTP in 20 µL 1 x PCR Buffer (Qiagen). After initial denaturation (98°C, 30 s) and 25 cycles of denaturation (98°C, 10 s), annealing (55°C 30 s) and elongation (72°C, 2 min), final elongation was carried out at 72°C for 5 min. PCR products were analyzed by agarose gel electrophoresis (1% wt/vol agarose), stained with ethidium bromide and visualized using a Kodak Image Station 2000 (Carestream Health).

2.6 Generation of pQE30TEV expression plasmid

Plasmid pQE30TEV, which codes for a TEV (Tobacco Etch Virus) protease recognition site C-terminal of a His_6-Tag (Fig. 9), was produced by PCR using vector pQE30 (Qiagen) as template, 1 unit of Phusion® High-Fidelity DNA Polymerase (New England Biolabs), 10

pmol of primers 1 and 2 (Table 1) and 0.04 mM of each dNTP in 50 µL 1 x Phusion® HF Buffer. After initial denaturation (98°C, 30 s), 26 cycles of denaturation (98°C, 10 s), annealing (55°C 30 s) and elongation (72°C, 3 min), final elongation was carried out at 72°C for 10 min. PCR products were analyzed as described earlier. After ligation with T4 ligase (Thermo Scientific) the plasmid was cloned in *E. coli* DH5-α (Invitrogen).

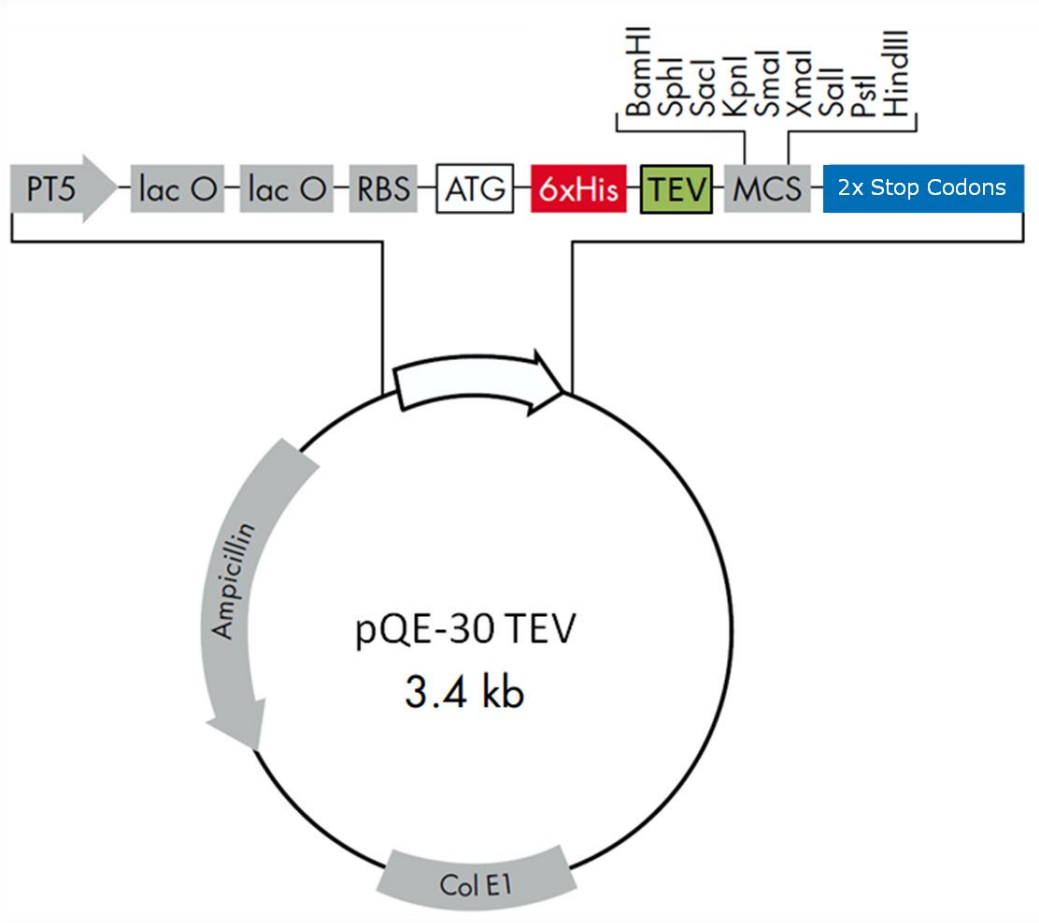


Figure 9: **Schematic representation of the pQE30TEV vector map.** The TEV coding sequence (green) was inserted by PCR in-frame to the histidine tag (red). Restriction sites are listed above the multicloning site (grey). The figure was adapted from Quiagen product details **(238)**.

2.7 Recombinant expression and isolation of proteins

Coding regions corresponding to the mature full length M proteins stG11, M1, M55, stG2078, M3.23, M3.0 and the M-like proteins Spa18 and Spa36 were amplified by PCR using specific primers 3 to 15 (Table 1) and the correspondent genomic DNA as template. The following PCR conditions were used: initial denaturation (98°C, 30 s), 25 cycles of denaturation (98°C, 10 s), annealing (55°C, 30 s) and amplicon specific elongation (72°C) (Table 1). The obtained PCR products were ligated between the *BamHI* and *SalI* cleavage site of the pQE30TEV vector. For overexpression of M proteins, M-like proteins and PARF-tandem, *E. coli* M15 pRep4 (Qiagen) were transformed with the produced plasmids and cultured in LB medium supplemented with ampicillin 100 µg/mL and kanamicin 25 µg/mL to an optical density of 0.5 at 600 nm wavelength. Then, protein production was induced by adding 1 mM IPTG (Isopropyl β-D-thiogalactopyranoside) and the bacteria were further cultivated overnight at 30°C. The bacteria were harvested by centrifugation (15 min; 18.000 × g), resuspended in LEW Buffer (Macherey-Nagel) containing protease inhibitor (cOmplete, Mini, EDTA-free, Roche) and homogenized using a French Press at 1000psi. Debris was removed from the lysate by centrifugation (20 min, 31.000 × g). Proteins were isolated from the lysate by affinity chromatography using Protino Ni-TED 2000 packed columns (Macherey-Nagel) following the manufacturer's protocol (Fig. 10).

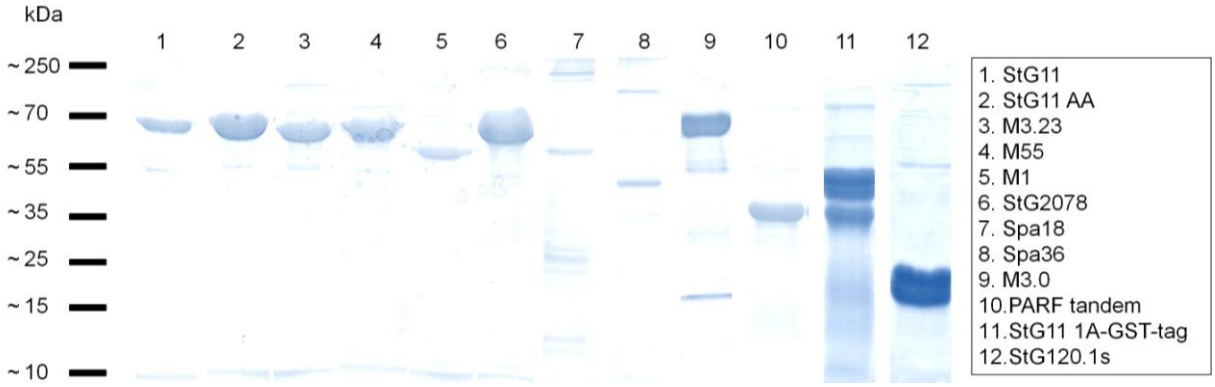


Figure 10: **Analysis of the affinity chromatography purification using a 12% SDS-PAGE.** A molecular weight standard is indicated on the left in kDa. The lanes are assignes to the recombinant proteins in the box on

the right. The amino acid sequences of the recombinant proteins are given in FASTA format in the supplements section.

Alternatively, the coding sequence of stG120.1 was used as template to amplify the truncated variants stG120.1s using primers 16 and 17, under conditions described in Table 1. After digestion with FastDigest™ EcoRV and subsequent ligation, the obtained plasmid pQE30TEV120.1s was cloned into *E. coli*. The protein StG120.1s that contained a His-tag was produced and purified as described above. StG11 1-A protein that contained a GST-tag was overexpressed in *E. coli* as described above and as in (**101**). The protein StG11 1-A was further purified by affinity chromatography using Glutathione-Sepharose 4B (GE Healthcare) and eluted with 10 mM reduced L-glutathione (Sigma-Aldrich) in 50 mM Tris-HCl (pH 8.0). All the isolated proteins were dyalized against PBS and stored at -20°C.

Reaction and Primers	Sequence 5'-3'	Elongation time [s]
Production pQE30TEV		
1	CACGGATCCGCATGCGAGC	180
2	GCGGATCCCTGAAAATACAGGTTTTCTCCGTGATGGTGA TGGTGATGC	
Amplification stG11		
3	GCGGATCCGCGGAGAATACATACGATAGATGG	60
4	GCTGTCGACTTATTAACCTGTTGATGGTAACTGTCTCTT	
Amplification AP-1		
5	GCGGATCCAACGGTGATGGTAATCCTAGGGAA	140
4	GCTGTCGACTTATTAACCTGTTGATGGTAACTGTCTCTT	
Amplification M3.0 or M3.23		
6	GCGGATCCGATGCTAGGAGTGTTAATGGAGAG	90
4	GCTGTCGACTTATTAACCTGTTGATGGTAACTGTCTCTT	
Amplification M55		
7	GGGGGATCCCAAACAGAACCATCTCAGACC	60
4	GCTGTCGACTTATTAACCTGTTGATGGTAACTGTCTCTT	
Amplification stG2078		
8	GAAGTTGAATCCGCGGAGAATACATAC	90
9	GACAAGATCATAATCTGCTTGTTT	
Mutation PCR stg11AA		
3	GCGGATCCGCGGAGAATACATACGATAGATGG	20
10	CCGAGCATATTCACCTATACTATTATTCTGC	
11	GGTGAATATGCTCGGGCTGCACAAAAACTTAATG	55
4	GCTGTCGACTTATTAACCTGTTGATGGTAACTGTCTCTT	
Amplification Spa18		
12	GCGGATCCGATTCAGTAAGTGGATTAGAG	60
13	GCTGTCGACTTATTACTCACCAGTAGATGGCAATTG	
Amplification Spa36		
14	GCGGATCCGAAGATAGGATAAATAGTGATATA	60
15	GCTGTCGACTTATTACTCACCAGTAGATGGTAATTG	
Truncation 120.1s		150
16	CCCGATATCTGAGTCGACCTGCAGCCAAGC	
17	GCGATATCTTATTAAACGGCAAGCTGATAAATTTC	
Colony PCR		
18	CCCGAAAAGTGCCACCTG	120
19	GTTCTGAGGTCATTACTGG	

Table 1: **Oligonucleotides used in this study.**

2.8 Site directed mutagenesis

The coding sequence of stG11 was mutated by a two-step PCR-based method using plasmid pQE30TEV-stG11 as template. The first step consisted of two PCR reactions. One amplified the DNA template from position 1 to 176 of the stG11 coding region using the primers 3 and 10 (Table 1). The other PCR introduced the mutation and amplified the fragment form position 163 to 1671 using primers 11 and 4 (Table 1). The two amplicons, purified with Nucleo Spin (Macherey-Nagel) and eluted in 30 µL NE Buffer, were mixed (2 µL of each amplicon solution) and used in the second step, an assembly PCR that amplified the mutated coding sequence of stG11AA (Fig. 11). In all three reactions, initial denaturation (98°C, 30 s) was followed by 25 cycles of denaturation (98°C, 10 s), annealing (55°C, 30 s) and amplicon specific elongation (72°C, Table 1). Elongation was carried out for 60 s in the assembly PCR. Final elongation was carried out at 72°C for 5 min. The PCR product was further digested with *BamHI* and *SalI* and cloned into the pQE30TEV vector as described above, yielding plasmid pQE30TEV-stG11AA. Subsequently, protein StG11AA was produced and isolated as described earlier.

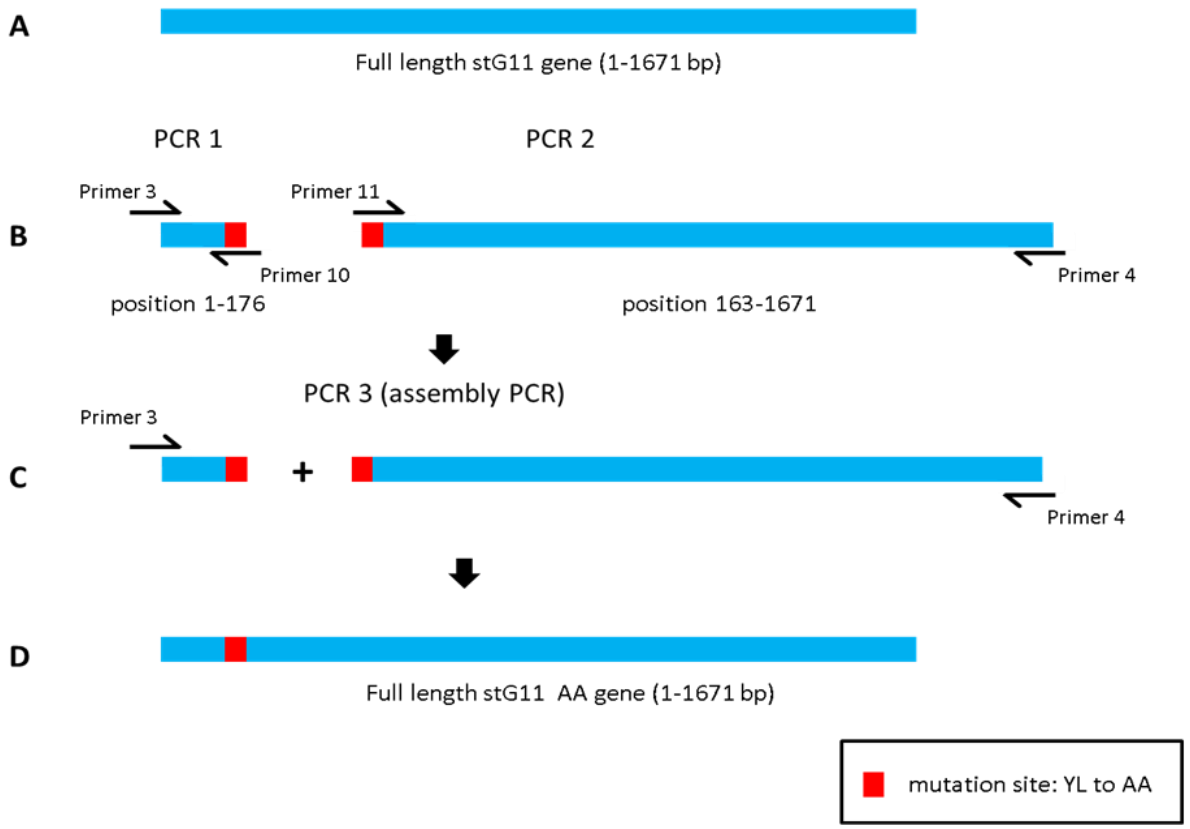


Figure 11: **Schematic representation of the site directed mutagenesis**. stG11 gene (**A**) was used as a template to perform the PCR 1 and 2 (**B**). An assembly PCR (**C**) was performed using the amplicons of the PCR 1 and 2 to introduce a substitution of four nucleotides in the coding region of the PARF motif of the stG11 gene (**D**). The site of the mutation YL to AA in the PARF motif is highlited in red.

2.9 Construction of a 20-valent PARF tandem coding sequence

A nucleotide sequence that codes for 20 different naturally occurring PARF sequences including three N- and four C-terminal flanking amino acids (PARF tandem, Fig. 12 A) was optimized for overexpression in *E. coli* using GenScript web tools and then chemically synthesized (supplements, table 2). The synthetic DNA was ligated into the vector pUC57 between the *BamHI* and the *SalI* restriction site in order to obtain the plasmid pUC57-PARF-tandem (GenScript, Piscataway, NJ). After multiplication of pUC57-PARF-tandem in *E. coli* DH5-α (Invitrogen) 0.1 ng of pUC57-PARF tandem was digested with *BamHI* and *SalI*. Subsequently, the PARF tandem coding sequence was cloned into the expression plasmid pQE30TEV between the *BamHI* and the *SalI* restriction sites. *E. coli* M15 pRep4 (Qiagen)

were transformed with the pQE30TEV-PARF-tandem and PARF-tandem was recombinantly expressed as described above (2.6, Fig. 12 B).

A

His$_6$-Tev [zzz(**A/T/E**)X**YL**XX(**L/F**)Nzzzz]$_{20}$

B

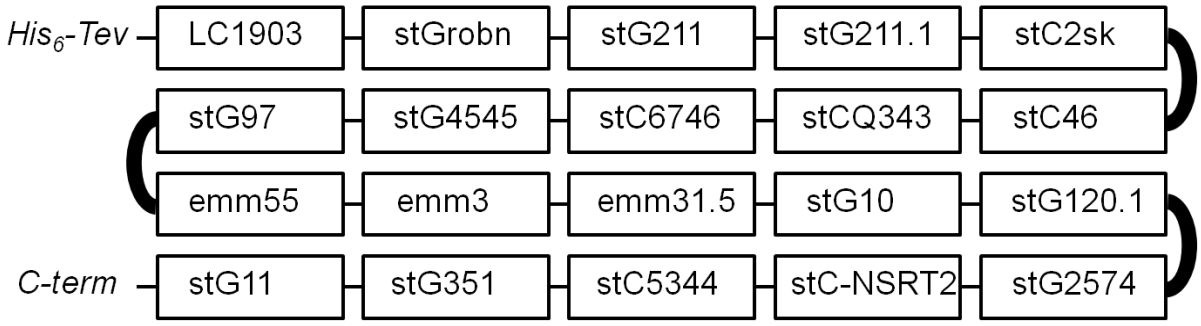


Figure 12: **Schematic diagram of the 20 valent PARF tandem protein. A.** The PARF motif consensus (conserved amino acids are given in bold) and flanking amino acids (z) are given in square brackets. **B**. *Emm*-types included in the PARF tandem construct are given in the boxes.

2.10 Labeling of Proteins with ^{125}Iodine

Labeling of proteins was performed as described by Hunter and Greenwood (**239**), using the chloramine-T method. This method is based on the iodisation of tyrosine residues of a given protein. Protein was dissolved in PBS to a final concentration of 1 mg/mL, and 100 µL of the suspension was transferred to a siliconized tube. 0.25 mCi carrier-free ^{125}iodine (Amersham) was added and the reaction was started by addition of 20 µL of chloramin-T (1 mg/mL in PB). After incubation at RT for 1 min, the reaction was stopped with 20 µL of sodiumdisulfite (1 mg/mL in PBS). The solution was filled-up to 2.5 mL with PBST and excessive radioactivity was removed by filtration using a PD-10 gel filtration column (Amersham). Radioactively labeled protein was eluted with 3.5 mL PBST.

2.11 Testing bacteria for collagen binding

Collagen I (from calf skin) and collagen IV (from human placenta) (Sigma-Aldrich) were labelled with radioactive iodine (^{125}I) (Hartmann Analytic). Streptococci were grown overnight (37°C, 5% CO_2) in TSB, washed and resuspended to a concentration of 10^8 bacteria per mL in PBST. Bacterial suspensions (250 µL) were incubated for 45 min with 10 µL solution of ^{125}I-labelled collagen (approximately 95 ng of ^{125}I-collagen I or 45 ng of ^{125}I-collagen IV, equaling 100.000 counts per minute). Unbound collagen was removed by washing in 1 mL PBST before radioactivity of the pellet was quantified in a Wallac 1470 Wizard gamma counter (PerkinElmer).

2.12 Ligand blots

Recombinant M proteins, PARF-tandem, (1, 0.5, 0.25, 0.125 µg) and Spa proteins (4, 2, 1, 0.5, 0.25 µg) were spotted onto nitrocellulose membranes (0.2 µm pores, Bio-Rad). After incubation for 1h at room temperature in PBST containing 5% skim milk and three steps of washing with PBST, the membranes were incubated overnight with radiolabeled collagen I (5.7 µg) or collagen IV (2.9 µg) in 10 mL PBST. Finally, the blots were washed thrice with PBST, before bound collagen was visualized using radiographic Amersham Hyperfilm ECL (GE Healthcare).

2.13 Biolayer interferometry

Bio-layer interferometry (BLI) is a label-free technology to measure interactions of biomolecules. BLI is an optical analytical technique that analyzes the interference pattern of white light reflected from two surfaces: a layer of immobilized molecules on the biosensor tip and an internal reference layer. The binding between the layer of immobilized molecules on the biosensor tip surface and an analyte in solution determine a wavelength shift that

correlates with the change in thickness of the layer. Thereby, interactions between molecules are measured in real time and the method allows to determine rates of association and dissociation and the dissociation constant with high precision (**240-241**).

Interactions between M proteins and collagen IV were measured using the BLI technology in an Octet® RED96 System (ForteBIO) at 25°C using PBS as running buffer. 2 mg of collagen IV in 1 mL PBS was biotinylated using EZ-link NHS-PEG12-Biotin (Thermo Scientific) following the manufacturer's protocol. The biotinylated collagen was purified using a PD-10 Desalting Column (GE Healthcare). Streptavidin coated biosensors (SA biosensors, ForteBIO) were pre-wet in PBS for 20 min. After a baseline step of 120 s in 200 µL PBS, the biosensors were transferred for 500 s into 200 µL solution of biotinylated collagen (30 µg/mL) for ligand binding, which resulted in immobilization of 0.5 to 0.8 nm of collagen. Control SA biosensors were treated the same, except that they were incubated in 200 µL PBS during the ligand binding step. Unspecific binding sites were saturated by incubating the biosensors for 200 s in 200 µL PBS containing 1% bovine serum albumin (BSA). Association of M protein with collagen was measured for 500 s in 200 µL solution of M protein in PBS, followed by the measurement of dissociation in 200 µL PBS for 200 s. The interactions were measured consecutively at M protein concentrations of 40, 10 and 5 µg/mL. Biosensors were regenerated between the measurements by washing for 30 s with 0.01% SDS in water. Data were processed and evaluated using the Octet User Software version 7.0 (ForteBIO). After subtraction of nonspecific binding, apparent dissociation constants (K_D) were calculated based on the Langmuir 1:1 binding model.

2.14 Synthetic PARF peptides

Previous work identified a synthetic PARF containing 15-mer peptide that was optimized for to collagen IV (peptide 17 GVY**RRYLETLND**RFQ) (**242**). For further experiment the

peptide 17 and a scrambled control (peptide 18 GVYRDLTLERYNRFQ) were produced by the in house service platform at the HZI. The peptides were synthetized using an Activo-P11 instrument and were coupled to an acetylated groupo in the N-terminal and an amide group in the C-terminal. The peptides were further purified with HPLC and stored at room temperature.

2.15 *In vivo* experiments

Permission for animal experiments was obtained from the Nds. Landesamt für Verbraucherschutz und Lebensmittelsicherheit (Az33.9-42502-05-11A170). Full-lenght M proteins and Spa18 were injected into the peritoneum of 7 weeks old female BALB/c mice. In groups of five mice, each mouse was given an emulsion of 25 µg M protein in 50 µL of sterile PBS (Amresco) and 50 µL of Freund's incomplete adjuvant (Sigma-Aldrich) per dose at days 1, 7, and 14. Mice of a control group received an emulsion of 50 µL of PBS and 50 µL of Freund's incomplete adjuvant. The animals were monitored for any symptoms of disorders. At day 21 the mice were sacrificed by CO_2 suffocation and serum samples were taken.

Alternatively, PARF-tandem antigen was injected into the peritoneum of ten 8-week-old female BALB/c mice following the same protocol of immunization used for Spa18 and M proteins.

For experiments with synthetic peptides, an emulsion of 50 µg of peptide 17 (GVY**RRYLETLND**RFQ) or peptide 18 (GVYRDLTLERYNRFQ) in 100 µL PBS and 100 µL Freund's incomplete adjuvant (Sigma-Aldrich) were injected into the peritoneum of 8 weeks old female BALB/c and C3H mice in groups of ten. Mice of a control group received an emulsion of 100 µL of PBS and 100 µL of Freund's incomplete adjuvant.

2.16 Human sera

Fifteen human sera, six from healthy donors and nine from ARF patients were described in a previous study (**113**). Four sera were recently collected from ARF patients in Chandigarh in accord with the local regulations. ARF was diagnosed applying the Jones criteria (**243**).

2.17 Enzyme-linked immunosorbent assay

Anti-collagen IV antibody titers in murine and human sera were determined as described previously (**101**) with few modifications. 96-well plates (flat bottom, high binding capacity, Greiner Bio-One) were coated overnight at 4°C with 50 µL of anti-human collagen IV rabbit polyclonal antibody (Progen, Heidelberg, Germany) diluted 1:100 in 0.1 M $NaHCO_3$ (pH 9.0) blocked with 2% BSA in PBST, and incubated with collagen IV (2 µg/mL in PBS) for 1 h at 37°C. After washing with PBST, 50 µL triplicates of diluted mouse sera (1:50, 1:150, 1:300 and 1:600), or human sera (1:50, 1:200, 1:800 and 1:3200) in PBS containing 1% BSA were added to the wells and incubated overnight at 4°C. After washing with PBST, bound antibodies were detected colorimetrically using horseradish peroxidase-coupled secondary antibodies goat anti-mouse IgG, Jackson Laboratories) or goat anti-human IgG Fc specific, (Sigma-Aldrich) and 2,2-azino-di-[3-ethylbenzthiazoline sulfonate] diammonium salt (ABTS, Roche Applied Science). The absorbance was measured at a wavelength of 405 nm in a Sunrise ELISA plate reader (Tecan).

Alternatively, 96-well plates (were coated over night at 4°C with 2 µg/mL of collagen I (calf skin, Sigma-Aldrich), laminin (human placenta, Sigma-Aldrich), or myosin (calcium activated, porcine heart, Sigma-Aldrich), respectively in 50 µL PBS. The wells were saturated by incubation with 2% BSA in PBST (1 h, 37°C). After washing with PBST, 50 µL triplicates of diluted mouse sera (1:50, 1:150, 1:300 and 1:600), or human sera (1:50, 1:200, 1:800 and 1:3200) in PBS containing 1% BSA were added to the wells and incubated overnight at 4°C.

Bound antibodies were detected colorimetrically as described above. Cutoff curves with a confidence level of 95% were determined based on control group data (**244**). The endpoint titer was as the interception of the regression line of the cutoff values and the regression line of the sample values, if the titer was lower than the reciprocal of the highest serum dilution tested. Higher titers were calculated as the interception between the regression line of the sample and the constant line determined by extrapolation of the cut of curve.

To determined the specificity of PARF-tandem antibodies in rabbit serum 96-well plates were coated over night at 4°C with 4 µg/mL of recombinant StG11 1-A or StG120.1s. Bound IgG was detected colorimetrically using horseradish peroxidase-coupled secondary antibody (goat anti-rabbit IgG, Jackson Laboratories) and ABTS. Cutoff curves were based on rabbit pre-immune serum.

2.18 Binding of Spa proteins to collagen IV

ELISA assay was performed to determine the binding between recombinant Spa proteins and collagen IV. A 96-well plate was coated overnight at 4°C with 5 µg/mL of collagen IV in 0.1 M $NaHCO_3$ (pH 9.0) and incubated for 1 h at 37°C with recombinant Spa18 or Spa36 (1 to 5 dilution series). After washing with PBST, the wells were saturated by incubation with 2% BSA in PBST (1 h at 37°C). After a further washing step with PBST, 50 µL triplicates of mouse serum (diluted 1:100) from animals injected with recombinant Spa18 was added to the wells and incubated overnight at 4°C. Bound IgG was detected colorimetrically using horseradish peroxidase-coupled secondary antibody as described above.

2.19 Histology

Heart tissues of the immunized mice were fixed in 3.7% formalin for 24 hours. Samples were further processed and examined at the Department of Pathology of the University of

Veterinary Medicine, Hannover, Germany. The samples were embedded in paraffin wax, sectioned and stained with hematoxylin and eosin for light microscopy.

2.20 Production of rabbit antiserum

Antiserum was produced by Davids Biotechnologie (Regensburg, Germany) by immunization of a rabbit with 100µg of PARF-tandem in a defined adjuvant 5 times, over a period of 60 days. Antiserum was collected two weeks after the final injection.

2.21 Bactericidal activity of PARF-tandem antiserum

The bactericidal activity of the rabbit antisera against *S. pyogenes* was measured as described previously with some modifications (**245-246**). Human neutrophils were isolated from blood of healthy volunteers using Polymorphprep (Axis-Shield, Norway) following the manufacturer's instructions. Isolated neutrophils were resuspended in RPMI medium (Sigma-Aldrich). Neutrophils (5 x 10^5 in 90 µL) were seeded per well of a 96 well plate. *S. pyogenes* strains stG11 (G45) and ΔstG11 (G89) were grown in THY medium to an optical density of 0.4 at 600 nm wave length. 50 µL of the bacterial culture was pre-incubated with 50 µL of anti-PARF tandem serum or pre-immune serum at 37°C for 20 min. Then 10 µL of the bacteria-serum mix was added to the neutrophils resulting in a multiplicity of infection (MOI) of 1 to 2 and incubated for 1 h at 37°C. Then, the bacteria were recovered from the 96-well plate and the number of colony forming unit (CFU) was determined by plating dilutions on blood agar plates. The bactericidal assay was repeated on three separate occasions.

To assess extracellular killing, neutrophils were pre-incubated with 1 µM cytochalasin D (Sigma-Aldrich) for 30 min at 37°C before adding streptococci.

CHAPTER 3

Results

3.1 Autoimmune responses in ARF patients

Sera of ARF patients were examined for presence of and possible interdependences between autoimmune responses in ARF. Samples of six healthy donors from the same geographic region were used to determine the cut-off curve for calculation of the titer. Eight out of 13 tested patients had a significantly elevated titer against at least one of the tested autoantigens: collagen I, collagen IV, myosin and laminin (Fig. 13). In four out of these eight patients, elevated autoimmune responses against collagens were concomitant with responses against myosin. Three out of these four patients had elevated titers against laminin, too. No link between the autoimmune responses was observed in the patients.

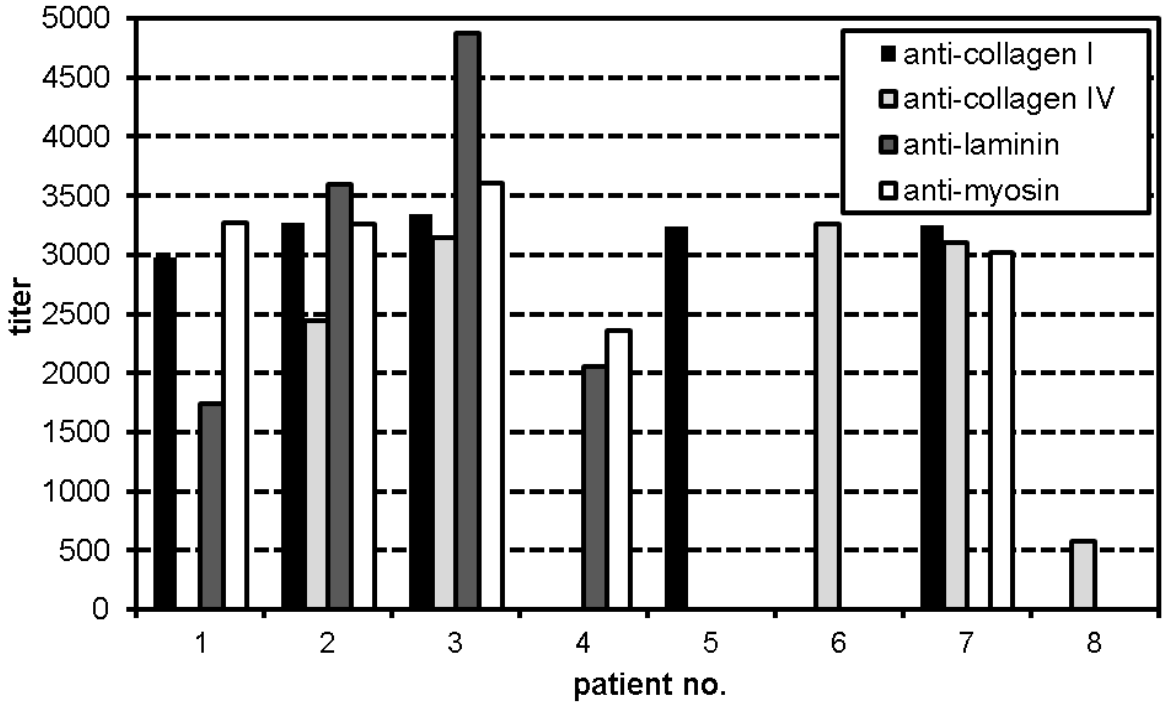


Figure 13: **Antibody responses against collagen I, IV, laminin and myosin in ARF patients.** A total of 13 sera of ARF patients were tested. Titers are shown for the eight out of 13 ARF patient sera that reacted at least against one out of the four tested autoantigens. The patient number is given on the x-axis.

3.2 Binding of collagen I and IV to *S. pyogenes* and SDSE

Streptococci bind collagen I and IV in a PARF-dependent (**59-101-102-247**) or in a PARF-independent manner (**113-196**). In a comparative experiment with streptococci of *emm*-types *emm*1, *emm*3.1, *emm*3.23, *emm*18, *emm*31, *emm*55, stG2078 and stG11, only isolates with a PARF-positive M protein and the highly encapsulated *emm*18 isolate bound radiolabelled collagen I and IV with high capacity (Fig. 14). Types *emm*1 and stG2078 that lack PARF and *emm*55 that bears an ineffective collagen binding motif (**59**) had the lowest binding capacity for both types of collagen. The PARF-negative isolate of type stG2078 bound significantly more collagen IV than other PARF-negative or M protein deficient isolates. Isolates of types *emm*1 and *emm*55 bound collagen at the same level as the M protein deficient mutants ΔM1 (Δ*emm*1) and G89 (ΔstG11) indicating that M1 and M55 protein did not contribute measurably to collagen binding. In absence of PARF or M protein, SDSE (G236, G89) bound more collagen I than *S. pyogenes* (A576, A799, A800, M1, ΔM1).

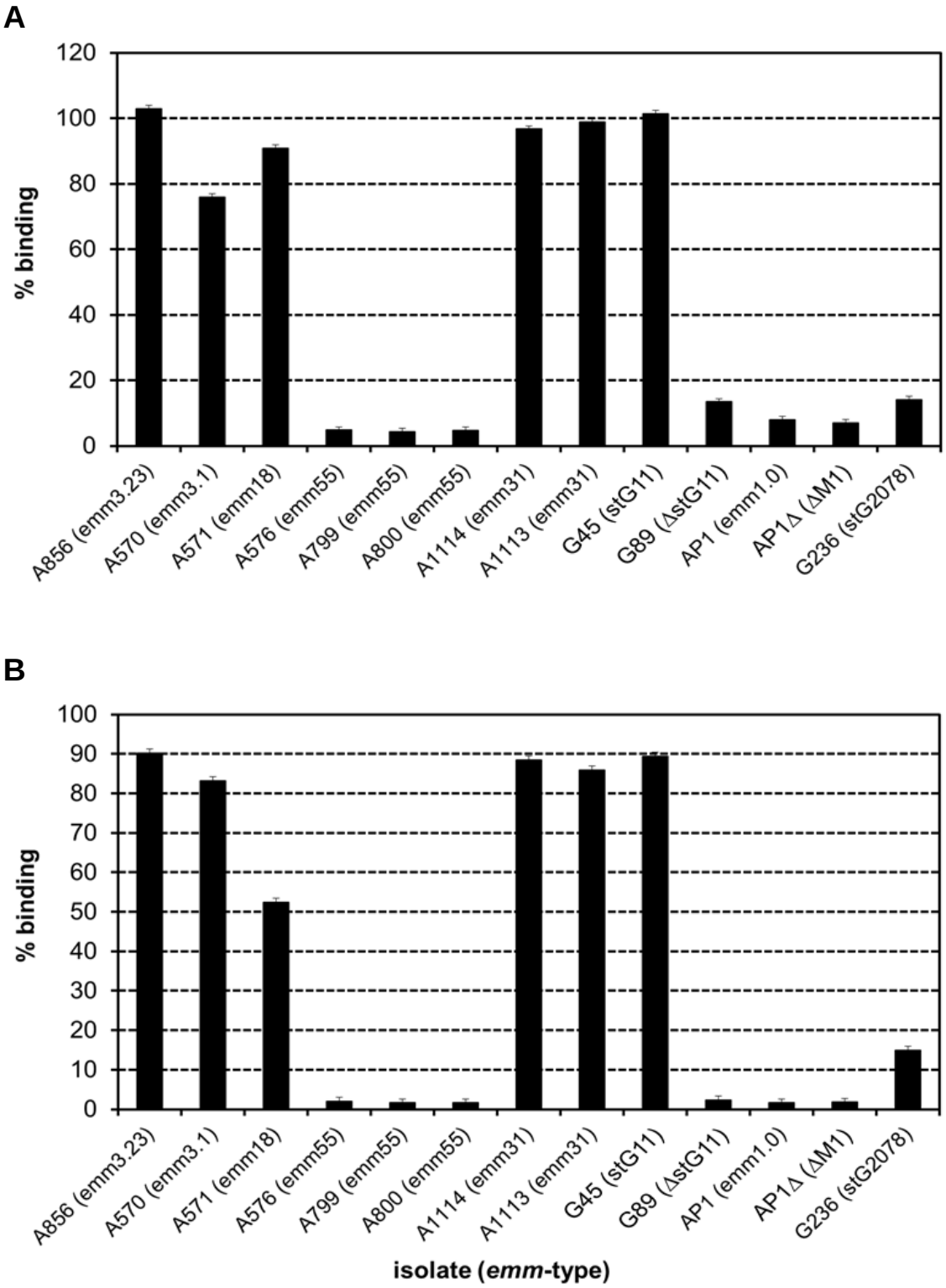


Figure 14: **Binding of collagen to *S. pyogenes* and SDSE.** The binding capacity for ^{125}I-labelled collagen I (**A**) and IV (**B**) was expressed as the bound percentage of added radiolabelled collagen (*black bars*) for isolates indicated on the x-axis. Error bars indicate the standard deciation of the triplicate measurements. All isolates belonged to the species *S. pyogenes* except of the SDSE isolates G45, G89 and G236. Isolates of types *emm*3.1, *emm*3.23, *emm*31, *emm*55 and stG11 carried M proteins with PARF motif.

3.3 Binding of collagen I and IV to M protein and Spa

Binding of collagen type I and IV was investigated in ligand blots with recombinant M proteins, including stG11AA, a variant of stG11 (also known as FOG) with inactivated PARF motif (ARAAQKLN) (**101**). Collagen I bound to 1 µg spotted M3.0 protein and, in a concentration dependent manner, to stG11 (Fig. 15).

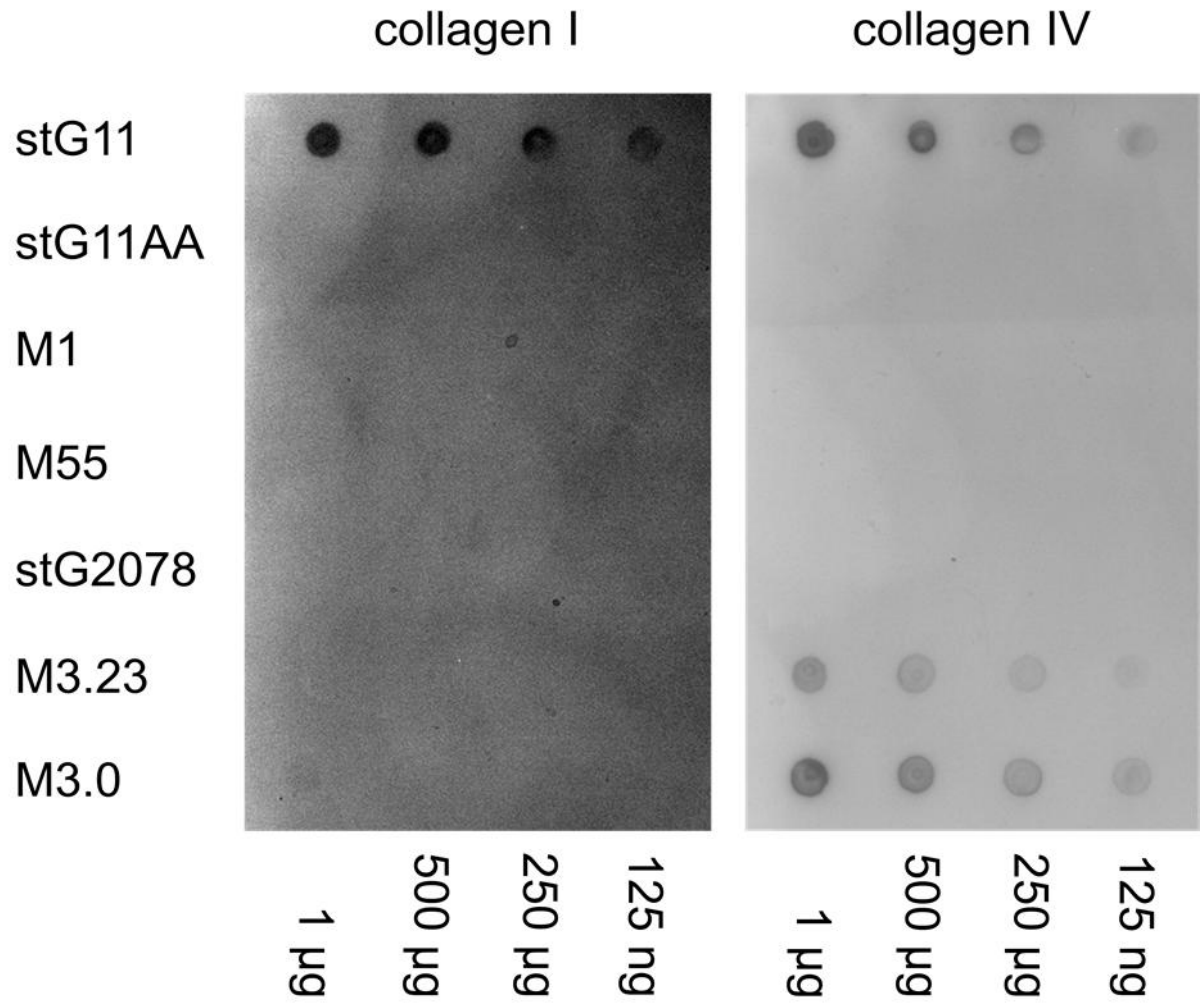


Figure 15: **Binding of collagen I and IV to M proteins in a ligand overlay assay.** Binding of radiolabelled collagen I (*left*) and collagen IV (*right*) to immobilized M proteins that are indicated on the *left*, was detected radiographically. Amounts of immobilized M protein are indicated below. Only a week signal for binding of collagen I to M3.0 at 1 µg could be detected.

Similarly, collagen IV interacted with stG11, M3.0 and M3.23 in a concentration dependent way, but with different affinity. As it was relevant for the interpretation of the immunization experiments that are presented below, the apparent dissociation constants for the interaction between collagen IV and stG11, M3.0 and M3.23 were determined by biolayer interferometry (BLI) as 15 nM, 52 nM and 160 nM, respectively (Fig 17). Notably, M3.23 differs from M3.0 in three amino acids, in position 31, 98 and 121 while PARF motif was located in positions 53 to 60 (Fig. 16).

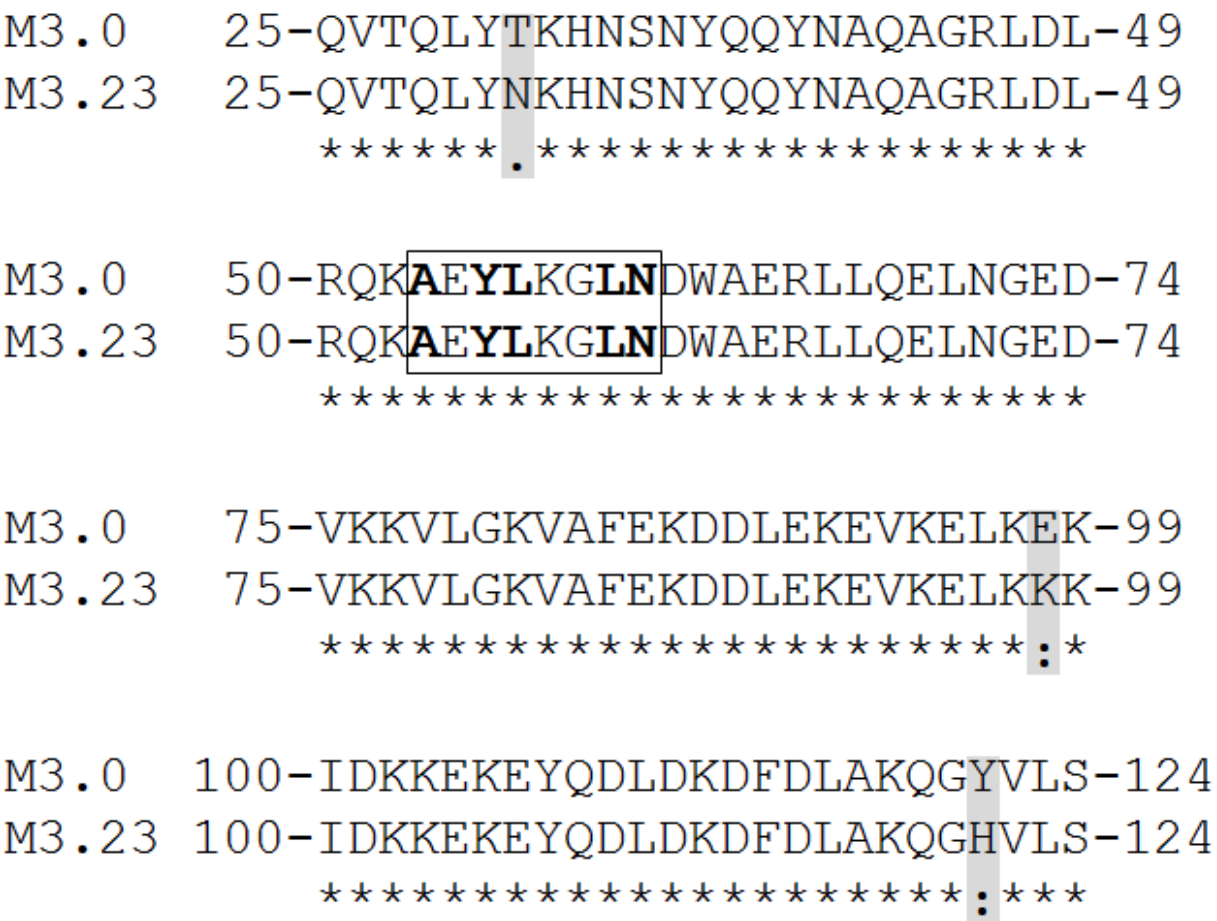
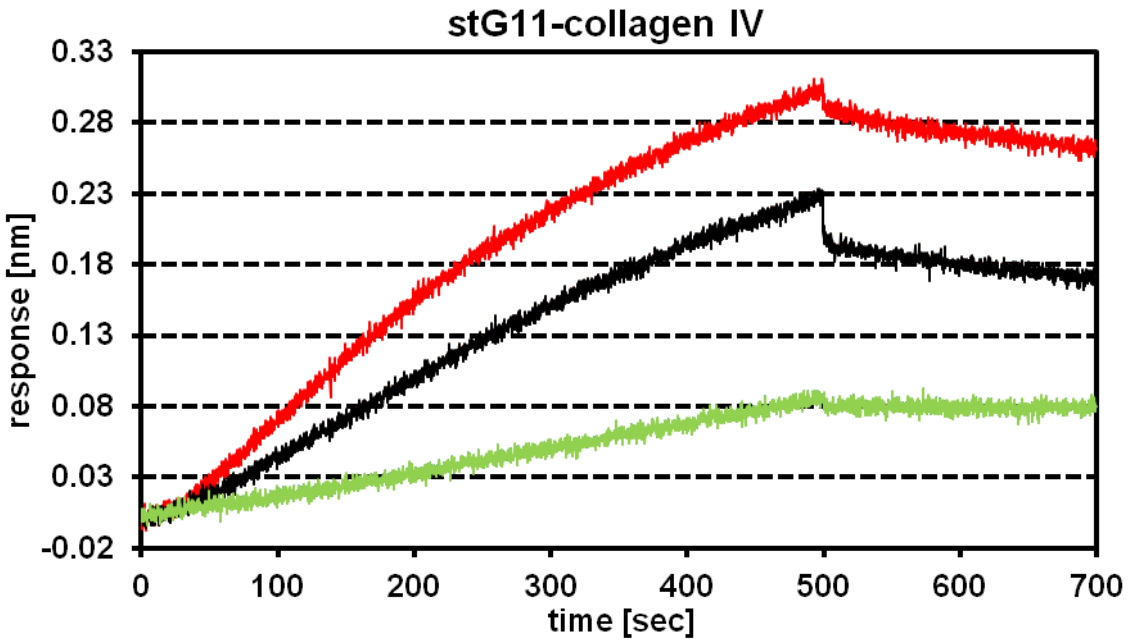


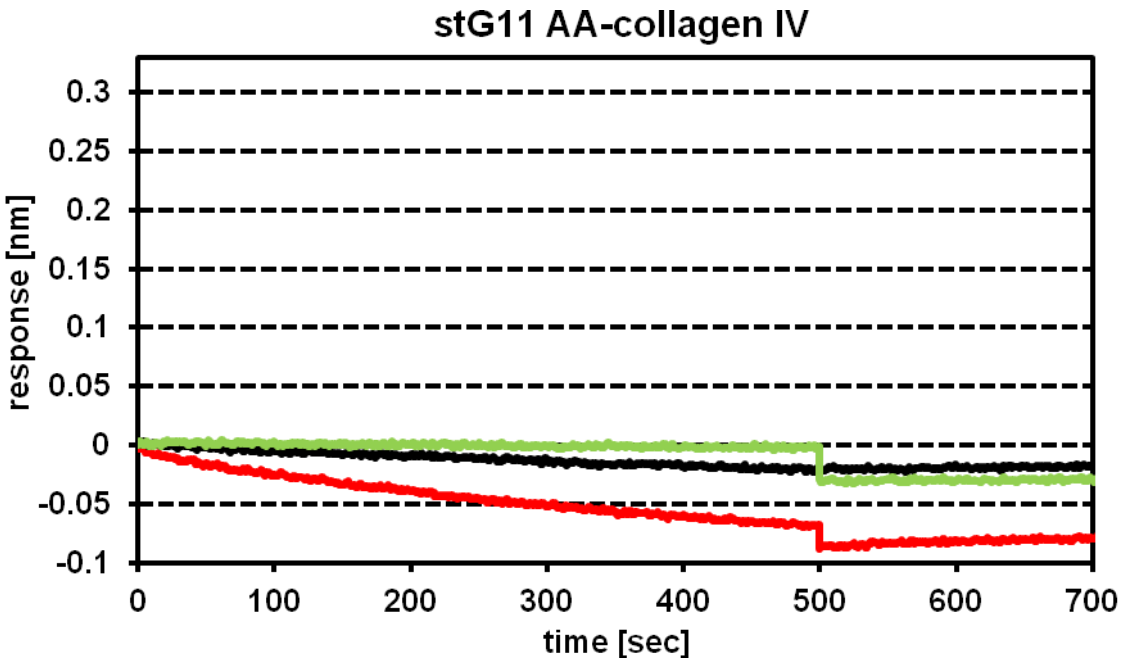
Figure 16: **Alignment of M proteins M3.0 and M3.23.** Partial protein sequences of two M3 protein sub-types M3.0 and M3.2 are depicted. The PARF motif is marked by a box. Characteristic amino acids of the PARF-consensus are indicated in bold. M protein designations are given on the left. Numbers that flank the sequences indicate their positions within the full length sequence of the mature protein without signal peptide. Substituted amino acids are highlighted in grey. Identical amino acids (*), conserved substitutions (:), and semi-conserved substitutions (.) are indicated below.

The decreased affinity of M3.23 for collagen IV, as compared to M3.0, indicates structural influences of other parts of the M protein other than PARF that impair collagen binding, as suggested previously (**247**). Despite lower affinity of M3.23, the corresponding *emm*3.23 isolate A856 had a high binding capacity for collagen IV. Potential reasons are a high avidity for collagen IV on the bacterial surface due to a high number of M3.23 molecules or additional or alternative collagen binding factors on A856. As in the ligand blot, no interaction was detected in BLI between collagen IV and M proteins stG11AA (Fig. 17 B), M1, M55 or stG2078. In accord with our previous observations and the binding experiments with bacteria that are described above, binding of M protein to collagen I and IV was limited to PARF-positive M proteins, but not all PARF-positive M proteins bound collagen with high affinity (**59-101-102**).

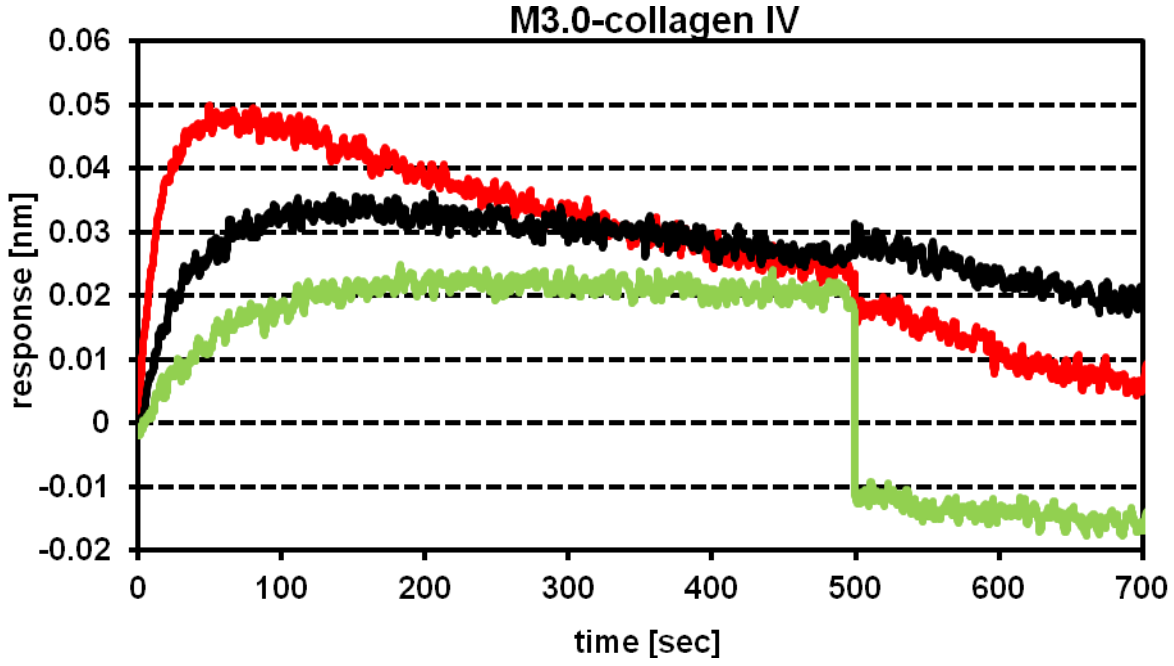
A



B


stG11 AA-collagen IV
response [nm]
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
0
100
200
300
400
500
600
700
time [sec]

C


M3.0-collagen IV
response [nm]
0.06
0.05
0.04
0.03
0.02
0.01
0
-0.01
-0.02
0
100
200
300
400
500
600
700
time [sec]

D

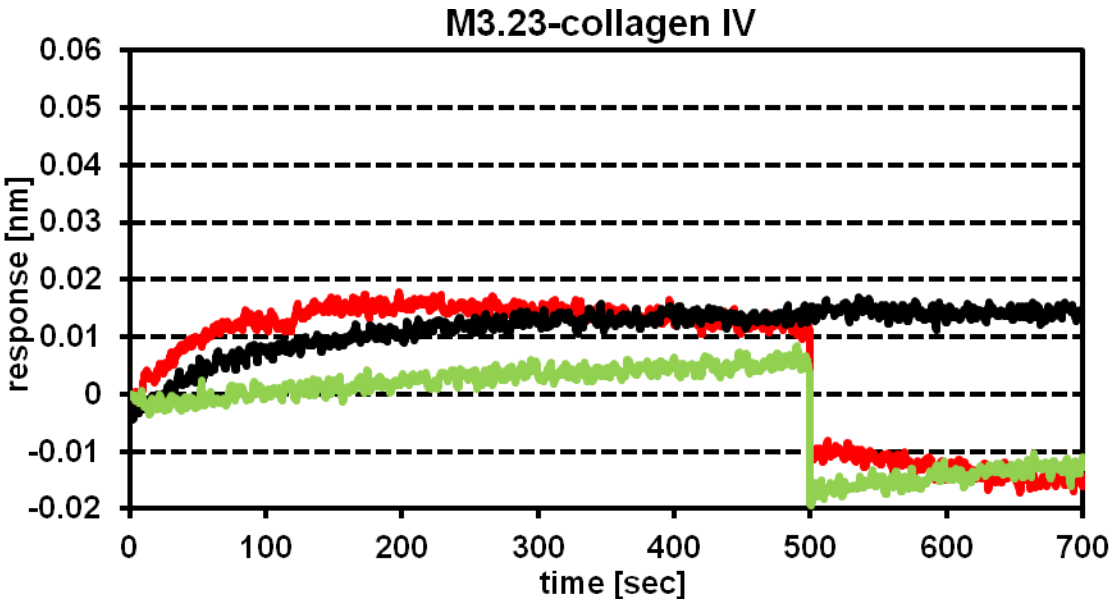


Figure 17: **Binding of M proteins to immobilized collagen IV in BLI measurements with SA biosensors.** Sensograms show the association (0-500 s) and dissociation (500-700 s) of the M proteins stG11 (**A**)**,** stG11 AA (**B**)**,** M3.0 (**C**)**,** M3.23 (**D**) in a concentration of 40 µg/mL (*red line*) 10 µg/mL (*black line*) and 5 µg/mL (*green line*)**.**

Binding of collagen I and IV to 4µg spotted Spa18 and Spa 36 was observed in a ligand blot experiment (Fig. 18). Collagen IV interacted with Spa18 and Spa36 in a concentration dependent way, but with different affinity.

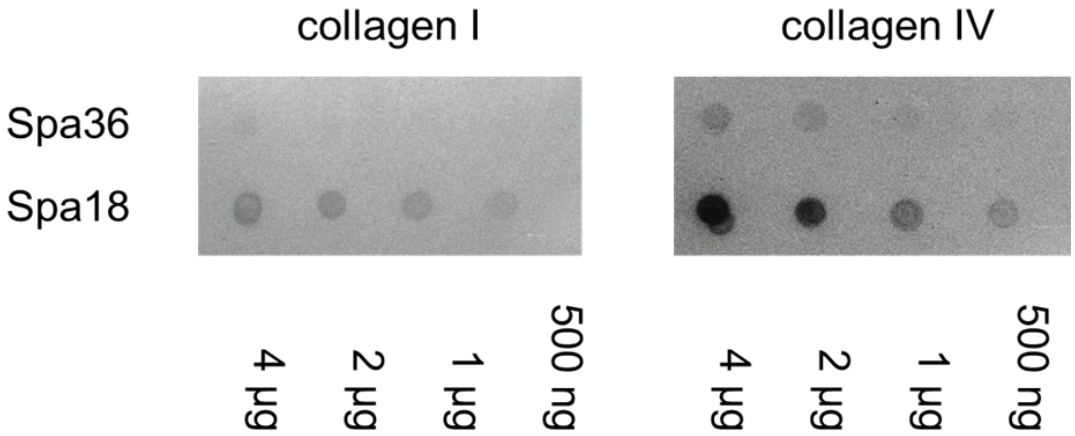


Figure 18: **Binding of collagen I and IV to Spa proteins in a ligand overlay assay.** Binding of radiolabelled collagen I (*left*) and collagen IV (*right*) to immobilized Spa proteins was detected radiographically. Amount of immobilized protein is indicated below. Only a week signal for binding of collagen I to Spa36 at 4 µg could be detected.

Binding of Spa proteins to collagen IV was confirmend by an ELISA experiment (Fig. 19).

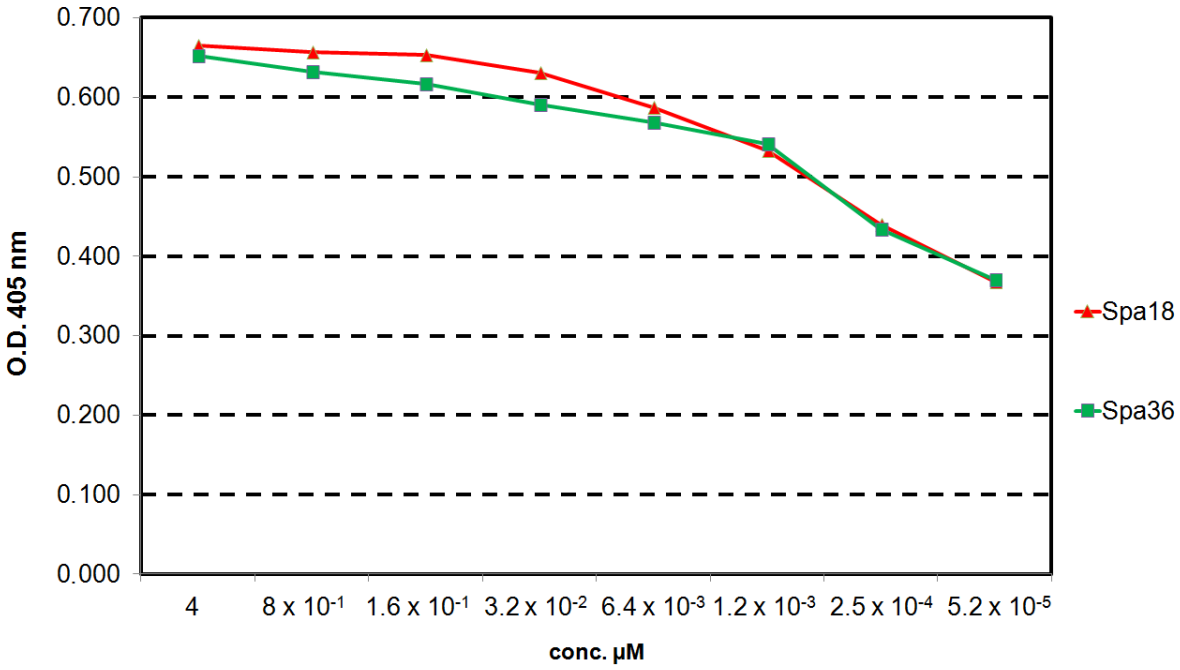


Figure 19: **Binding of Spa proteins to 96-well plate coated with collagen IV.** Concentration of recombinant Spa18 (*red line*) and Spa36 (*green line*) is given on the x-axis (µM), value of optical density is given on the y-axis.

3.4 Induction of autoantibody production by M proteins and Spa18 *in vivo*

Recombinant M proteins that differed in affinity for collagen IV were examined for their ability to induce the production of antibodies against collagen IV, myosin and laminin in mice. Reactivity against these autoantigens has been implicated in ARF pathogenesis (**101-113-174**). In addition, antibody titer against collagen I was determined and it was sought for causative connections between the investigated autoimmune responses (Fig. 20). Heart of mice that mounted autoimmune responses against collagen IV and other autoantigens were devoid of histopathological abnormalities in heart musculature and valves.

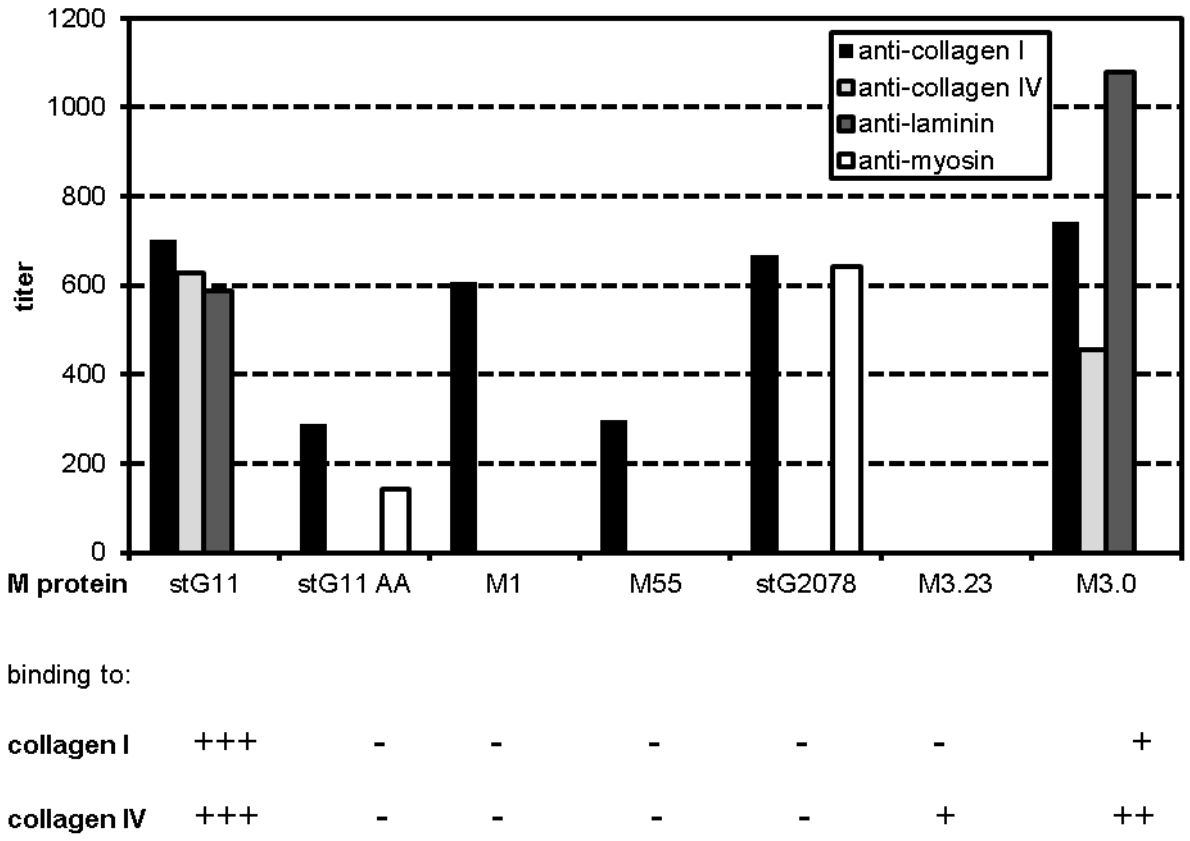


Figure 20: **Antibody responses against collagen I, IV, laminin and myosin induced in mice by M proteins.** Mean titers are shown for each group of mice, which were immunized with the M protein that is given on the x-axis. The ability of the respective M protein to bind collagen I or IV is indicated below. +++ high affinity, ++ intermediate affinity, + low affinity, - no binding detected.

Elevated response against myosin arose in the groups of mice that were immunized with stG11 AA (titer = 143) and stG2078 (titer = 643). Anti-myosin response that occurred in these mice did not coincide with an anti-laminin response, which was observed in the groups of mice that received stG11 or M3.0. The results indicate no or low crossreactivity between laminin and myosin autoantibodies in the utilized *in vivo* model for induction of autoimmunity by M protein.

All the M proteins induced collagen I autoimmunity in mice, except of M3.23. This induction of autoimmunity did not depend on the presence of a functional PARF motif because it also

occurred in the groups of mice that were injected with stG11AA, M1 or stG2078, which lack a PARF motif.

In contrast, the experiments unambiguously demonstrated that collagen IV autoimmunity, caused by M protein, depends on a fuctional PARF motif. Antibody titers against collagen IV increased significantly only upon injection of stG11 (titer = 629) or M3.0 (titer = 456), which both bind collagen IV with high affinity via a PARF motif. Injection of M55 or M3.23, which have a lower affinity for collagen IV as compared to stG11 or M3.0 (Fig. 20), did not evoke collagen IV autoimmunity, indicating that PARF-positive M proteins require a certain affinity for collagen IV to trigger this autoimmune response (Fig. 20).

Interdependences between anti-collagen and anti-myosin responses could be excluded (Fig. 20). None of mice that developed autoimmunity against myosin showed a significant response against collagen IV, excluding that collagen IV autoimmunity is a side-effect of the previously described molecular mimicry of myosin (**174**).

Notably, elevated titers against laminin occurred exclusively in the groups of mice that were injected with M proteins that triggered collagen IV autoimmunity, M3.0 (titer = 1078) and stG11 (titer = 587). This suggests that collagen IV autoimmunity preceded laminin-specific responses in this *in vivo* model rather than the reverse. However, an experiment with Spa18 does not suggest a causative connection between these two responses (Fig. 21).

Taken together, the experiments prove that the collagen binding PARF motif is a specific trigger of collagen IV autoimmunity *in vivo*.

Spa is a M-like protein that is predicted to form a coiled coil structure and is located on the surface of *S. pyogenes* of *emm*-types 18 (Spa18) and 36 (Spa36). Spa18 and Spa36 did not harbour a PARF motif but bound to collagen IV (3.3, Fig. 17 and 18). As *S. pyogenes* of *emm*-type 18 caused several ARF outbreaks in the USA (**142**) Spa18 was investigated in our

established *in vivo* model as a possible trigger of collagen IV autoimmunity and of the other ARF-related autoimmuneresponses (3.1, Fig. 13). In particular, Spa18 was examined for its ability to induce the production of antibodies against collagen I, collagen IV, myosin and laminin in mice (Fig. 21).

Notably, Spa18 trigger autoimmunity against collagen I (titer = 303), collagen IV (titer = 641) and myosin (titer = 624), but not anti-laminin response.

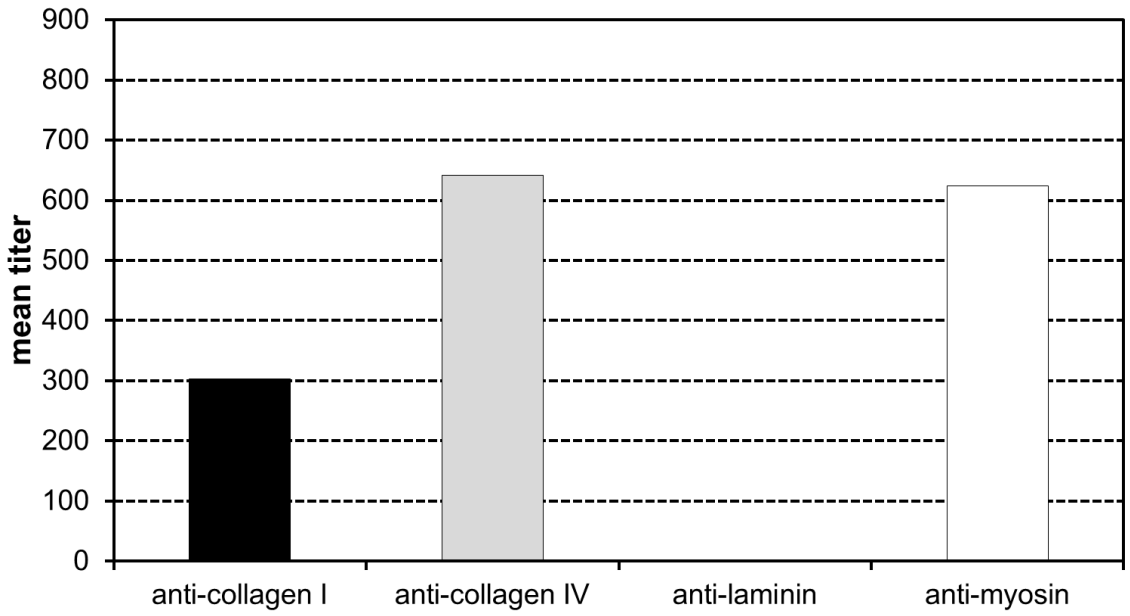


Figure 21: **Antibody responses against collagen I, IV, laminin and myosin induced in mice by Spa18 protein**. Mean titers for the group of immunized mice are given on the y-axis.

The lack of correlation between the investigated autoimmune responses in the experiments with M proteins and Spa is in accordance with the responses that were observed in ARF patients (3.1). A causative connection beween anti-collagen and anti-coiled-coil responses is not suggested by the presented data.

3.5 Binding of collagen IV to PARF tandem protein

PARF tandem consist of all the PARF motif known to date. This makes the recombinant PARF tandem a valuable tool to develop a vaccine that protect against all the *S. pyogenes* and SDSE that harbour PARF-containing M protein. Several experiments show that M protein that harbour PARF motif and synthetic PARF peptides bind to collagen IV with different affinity. However, the binding of PARF tandem protein to collagen types needed to be examined.

Binding of PARF tandem to collagen type IV was investigated in ligand blots (Fig. 22). Collagen IV bound to the spotted PARF tandem protein in a concentration dependent manner, however, the PARF tandem construct did not bind to collagen I in ligand blot assay (Fig. 22).

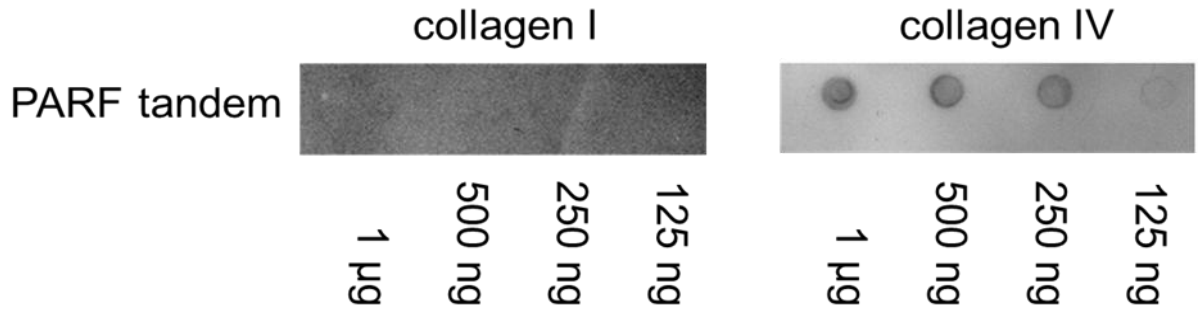


Figure 22: **Binding of collagen I and IV to PARF-tandem protein in a ligand overlay assay.** Binding of radiolabelled collagen I (*left*) and collagen IV (*right*) to immobilized PARF-tandem protein was detected radiographically. The amount of immobilized protein is indicated below.

3.6 Immunity against PARF tandem and its reactivity against PARF containing M proteins

PARF tandem antigen is a potent inducer of specific antibodies in rabbit (Fig. 23). Moreover, the serum reacted against the N terminal part of StG11 (named StG11 1-A) and StG120.1 (named StgG120.1s), which both carry a PARF motif. The experiment was carried out with N-terminal fragments, which, in contrast to the full length M proteins, do not bind IgG by the Fc-region (**248**). The result indicates that PARF tandem evokes immune responses against PARF containing M proteins, presumably due to antibodies that react against PARF. The exact specificity of the antibodies requires further investigation.

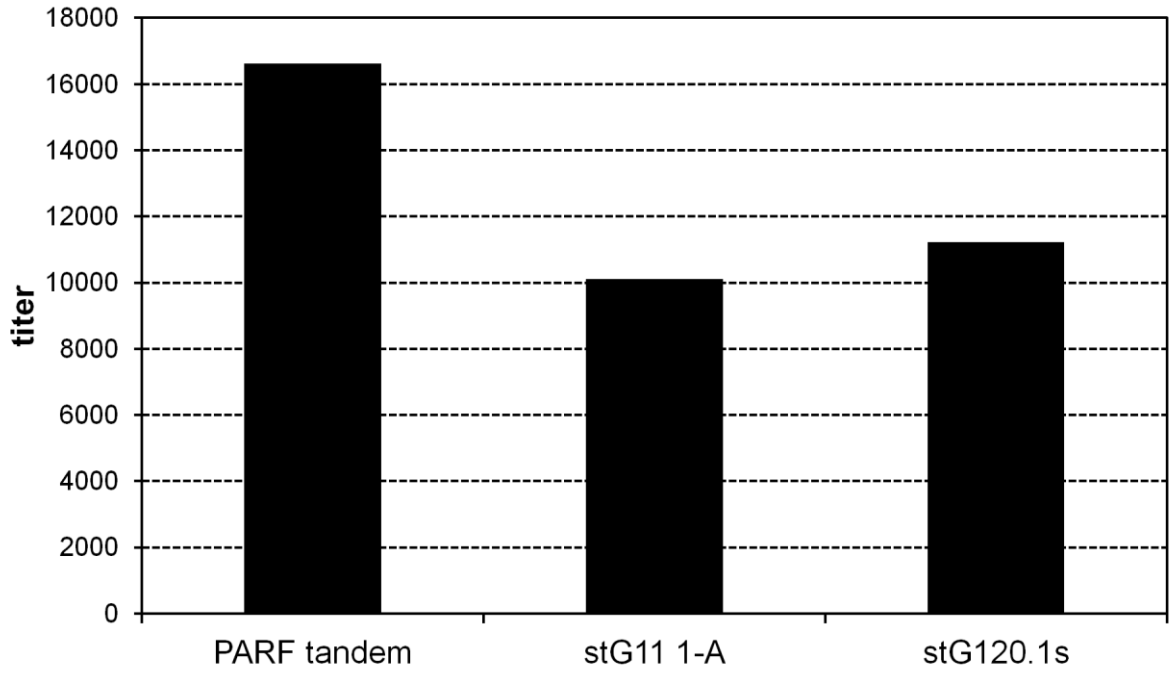


Figure 23: **Reactivity of PARF tandem antiserum that was raised in rabbit.** The titers of rabbit antibodies against PARF tandem and the N-terminal fragment of M protein stG11 and stG120.1s are given on the y-axis.

3.7 Opsonizing activity of antibodies against PARF tandem

To determine the opsonizing activity of the antibodies that were produced in response to immunization with the PARF tandem antigen, an *in vitro* test was performed using the strain stG11, that harbors a M protein with a PARF motif (StG11) and its natural mutant, the strain ΔstG11 that does not express M protein on its surface. The bactericidal assay determined the percentage of streptococci that survived exposure to human neutrophils after pre-incubation with either pre- or immune rabbit serum. Immune rabbit serum was opsonizing and bactericidal just for strain stG11 (37% of survival), indicating that the bactericidal activity of the serum depended largely on M protein specific opsonization (Fig. 24).

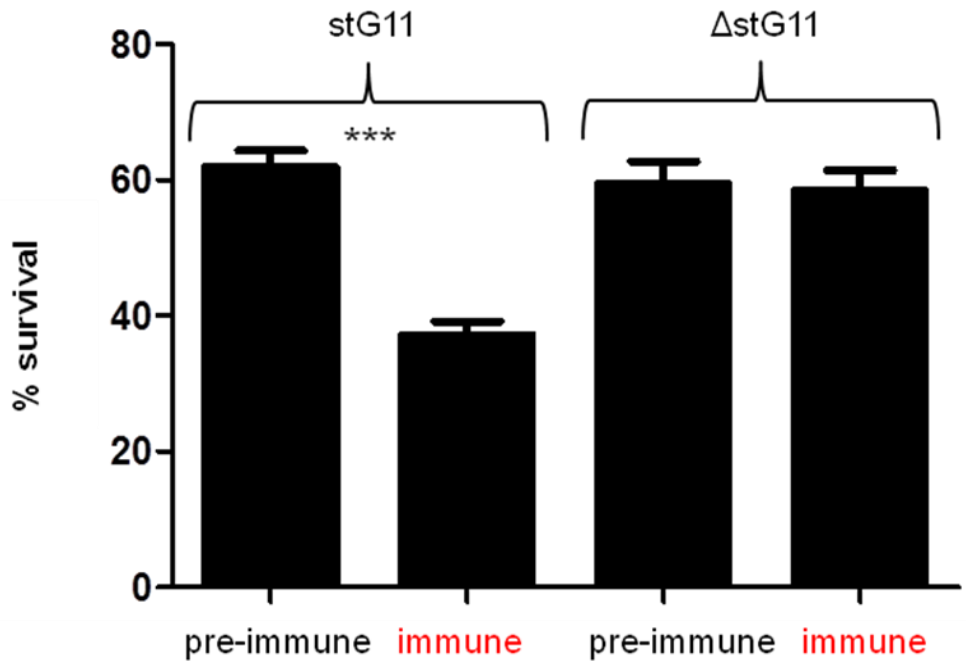


Figure 24: **Effect of PARF tandem antiserum on the bactericidal effect of human neutrophils *in vitro).*** Survival rate of SDSE stG11 that expressed a PARF positive M protein and its isogenic mutant ΔstG11 that lacked M protein on its surface is indicated on the y-axis. As indicated on the x-axis, the bacteria were pre-incubated either in serum of a rabbit that was immunized with PARF (immune) or with the pre-immune serum of the same animal. The survival of stG11 preincdubated in rabbit sera is significantly reduced (p value < 0.0001, ***) when compared to stG11 non incubated in rabbit sera. Experiment performed by Dr. Uchiyama, HZI Braunschweig.

3.8 Induction of autoantibody production by peptide 17 and PARF-tandem *in vivo*

The synthetic PARF peptide (peptide 17) and its scrambled control (peptide 18) were examined for their ability to induce the production of autoantibodies against collagen IV in mice. Injection of peptides 17 and 18 did not evoke collagen IV autoimmunity, suggesting that the small PARF containing peptide 17 may not be sufficient to induce autoimmuninity in our animal model. Secondary structure prevision indicated that peptide 17 do not form coil-coiled, a property of full-lenght M protein. This characteristic may be crucial to explain the the lack of autoimmunity against the peptide 17. In fact, although the affinity to collagen IV for peptite 17 is moderate, the lack of seconday structures may reduce its avidity for collagen and decrease the adherence properties that the PARF motif confers to M proteins.

Notably, antibody titers against collagen IV (titer = 307) and myosin (titer = 804) increased significantly only upon injection of PARF tandem (Fig. 25). In contrast, PARF-tandem did not trigger autoimmunity against collagen I or laminin in the mice.

Elevated response against collagen IV or myosin was concomitant in 7 out of 10 mice injected with PARF tandem. However, a causative connection between anti-collagen IV and anti-myosin response was excluded by the experiment with full-lenght M proteins.

PARF-tandem lacked heptad repeat structures. Therefore, induction of myosin autoimmunity by PARF tandem did not depend on the presence of a coiled-coil structure. This suggests that molecular mimicry was either independent of the coiled-coil structure or the autoimmune response against myosin was caused by an alternative mechanism. Binding of myosin to PARF-tandem remains to be tested.

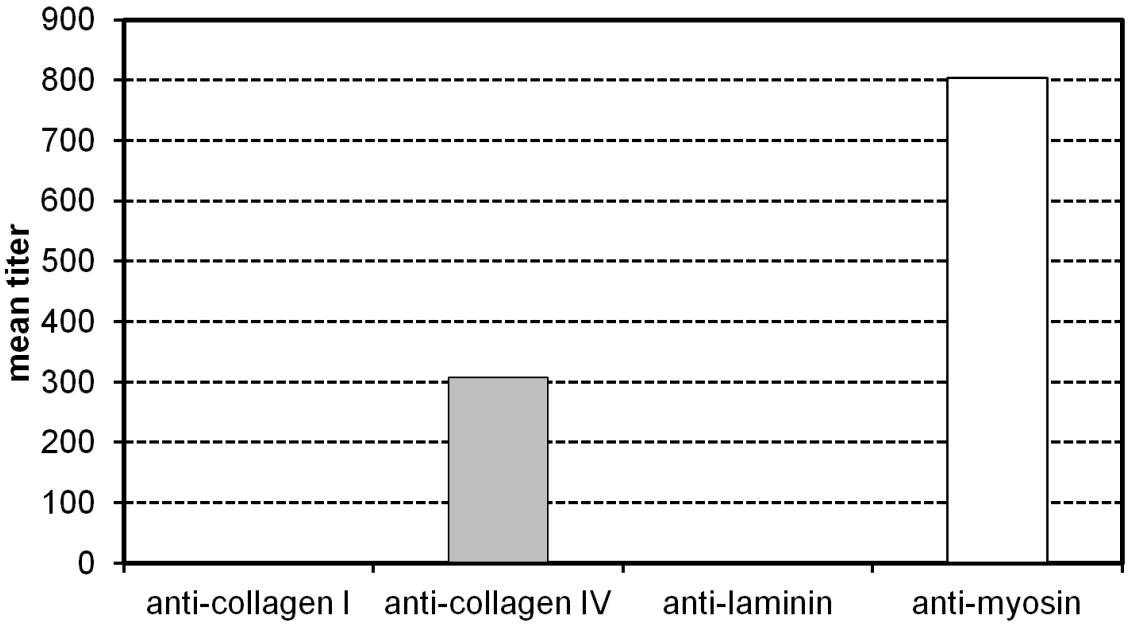


Figure 25: **Antibody responses against collagen I, IV, laminin or myosin induced in mice by PARF tandem protein.** Mean titers for the group of immunized mice are given on the y-axis. Specificity of the antibodies is given on the x-axis.

The injection of PARF tandem led to antibody responses against collagen IV and myosin in the animal model. This is further evidence for the role of PARF as the trigger of autoimmunity. In addition, the result indicates that PARF containing antigens when used for immunization are associated with a risk of side effects.

CHAPTER 4

General discussion

4.1 General discussion

Destructive autoimmune responses in ARF are limited to the extracellular matrix (ECM), as indicated by histopathological examinations on patient samples (**166**). This includes rheumatic carditis, which is a frequent manifestation of ARF and the suspected cause of RHD. Histopathologically, rheumatic carditis presents as an interstitial carditis, rarely showing signs of myocarditis. Moreover, markers for myocardial damage, such as troponins and creatine kinase, are not significantly elevated in these patients (**166-249**). Since cardiac myosin is localized inside the myocyte, these observations suggest that this host protein is not a primary target for the poly-reactive autoimmune responses in ARF (**65-166-174**). Instead, they could react with coiled-coil proteins in the ECM, basement membranes or on the cell surface. One potential target is laminin that forms a triple-helical coiled-coil and is recognized by monoclonal antibodies from patients with rheumatic carditis (**250**). Like collagen IV, laminin is a major component of basement membranes (**251**), including the basement membranes that support the endothelium of heart valves. Therefore, autoimmunity against these proteins may be directly involved in the degeneration of the heart valves in RHD.

Notably, reactivity against all the tested autoantigens collagen I and IV, laminin and myosin was observed in ARF patients (Fig. 13). However, no correlations between these responses were observed. A low degree of cross-reactivity between anti-myosin and anti-laminin responses was noted, which indicates that the triggering factors were not limited to previously identified molecular mimicry epitopes that lead to the production of poly-reactive antibodies.

An association of anti-laminin responses with M proteins that triggered collagen IV autoimmunity via a PARF motif (Fig. 20) was not supported by the experiment with PARF tandem (Fig. 25). Moreover, PARF-independent induction of collagen IV autoimmunity by

Spa18 was not concomitant with anti-laminin reactivity (Fig. 21), which also speaks against a causative connection between these responses.

After injection of M protein, mice that mounted an autoimmune response against collagen IV and other autoantigens did not develop relevant pathologic changes in cardiac tissue during the short period of the experiment. Thus, the utilized *in vivo* model was restricted to processes that are likely to precede manifestation of ARF. Consequently, it is unlikely that the observed immune responses are not secondarily caused by tissue damage.

The ability of M proteins to trigger collagen IV autoimmunity is confirmed, associated with the N-terminal part of the protein (**101**) and unambiguously linked to the PARF motif (Fig. 20). The underlying molecular and cellular mechanisms of PARF-dependent autoimmunity are not yet explored, but appear to involve collagen binding. Moreover, the causes for the autoimmune responses against collagen I and in part of laminin and myosin remain elusive. Autoimmunity against collagen I was not associated with the function of PARF (Fig. 20 and 25). Induction of immunresponses against laminin was limited to M proteins that bound collagen with high affinity (Fig. 20), thus may involve an alternative mechanism to molecular mimicry. Similarly, myosin autoimmunity was triggered by PARF tandem that lacks the heptad repeads for coiled coil structures, which could cause molecular mimicry (Fig. 25). Further research on the causes of these responses is needed.

PARF is an epidemiologically relevant collagen binding motif that occurs in about 7% of the *S. pyogenes* isolates from human infections. Moreover, the discovery of Spa as a novel collagen binding protein that triggers autoimmunity against collagen IV, despite lacking a PARF motif, broadens the spectrum of potential rheumatogenic factors in *S. pyogenes*. Spa18 is the first trigger of autoimmunity that was identified in *emm*-type18, which caused several outbreaks of ARF in the USA (**72**-**142**-**182**). Some *emm*1 and *emm*18 isolates are

encapsulated by hyaluronic acid and this capsule has been suspected as the collagen binding factor of *emm*18 isolates. (**113**) However, an ability of hyaluronic acid to trigger collagen autoimmunity has not been shown.

Only *emm*18 and PARF-positive *S. pyogenes* or SDSE bound collagen I or IV with high capacity in this or previous studies (**102-113-250**). Previously reported binding of collagen I or IV to *emm*1 (**196**) was not detectable in comparative binding experiments neither with whole bacteria nor with recombinant M protein (Fig. 14 and 15). Moreover, a previous experiment excluded an interaction between the PARF-negative protein M18 and collagen IV (**113**). Thus, hyaluronic acid (**113**) or other factors such as Slr (**196**), Cpa (**195**) or Spa18 (**141**) may be employed for collagen binding by PARF-negative *S. pyogenes.* So far, an ability other factors than M proteins to trigger autoimmunity is only shown for Spa18.

The utilized *in vivo* model identified early antibody responses to M protein and showed that PARF and molecular mimicry are independent triggers of autoimmunity. In ARF, additional streptococcal factors, individual conditions in the host and progression of the pathogenesis may have contributed to the diversity of autoimmune responses that was observed between different patients (Fig. 13). In four out of the total of 13 examined patient sera, no reactivity against any of the examined autoantigens was detected. The reactivity may have ceased during the course of diseases or other than the examined autoantigens may have been involved in these cases. By causing inflammation, tissue damage and epitope spreading, initial autoimmunity against one or a few proteins of the extracellular matrix may initiate a cascade of autoimmune responses that lead to the various symptoms of ARF. Only seemingly, this is contradictive to the specificity of the degenerative processes for the heart valve in subsequent RHD, which may ground on a limited ability of this tissue to heal and to resolve the destructive autoimmune responses, rather than on the specificity of the autoimmune responses (**166**). In this concept of ARF pathogenesis, single autoimmune responses caused by PARF or

molecular mimicry have a triggering function. Occurring in 8 out of 13 patients they may be valuable diagnostic and prognostic markers for ARF and RHD. As the responses in the *in vivo* model, the responses in patients indicate that PARF-induced autoimmunity and molecular mimicry are two independent triggering pathomechanisms of ARF, but they may act synergistically in some of the patients as they have been co-detected in four out of the 13 patients.

For the first time we proved that the PARF motif is an antigenic part of M protein and injection of PARF tandem in animals produced *emm*-type specific antibodies (Fig. 23). These antibodies may be a useful tool for detection of rheumatogenic streptococci, An antibody-based test would improve diagnosis by identifying patients with an increased risk to fall ill with ARF or RHD.

PARF tandem evoked the production of opsonizing antibodies in rabbit (Fig. 24) and could be an interesting vaccine candidate. Including PARF in vaccines may protect against the infection caused by PARF-positive *S. pyogenes* (*emm*3, *emm*31 and *emm*55) but also SDSE (Supplement D) broadning the coverage of a streptococcal vaccine by targeting isolates with the potential to cause collagen IV autoimmunity. However, the *in vivo* experiments indicated the possibility of immune side-effects (Fig. 25), thus, the 20-valent PARF tandem in its present form may be not safe for use in humans. A PARF-based or PARF-containing vaccine would require a careful redesign.

In summary, the presented data provide evidence that collagen IV autoimmunity is specifically induced by PARF. Thus, it is independent of other host responses to M protein. However, it requires further investigations to elucidate the role of collagen IV binding in this barely understood immune response. Other streptococcal factors, as shown for Spa18, trigger autoimmunity in a similar way. A better understanding of the underlying processes will help

to conceive preventive measures against ARF. Moreover, it may also increase our understanding of other rheumatic diseases that may originate from noxious host-microbe-interactions.

REFERENCES

1. Hollricher K. (2007). Species Don‘t Really Mean Anything in the Bacterial World. Lab Times. 5 p23 ff.
2. Köhler W. (2007). The present state of species within the genera Streptococcus and Enterococcus. Int J Med Microbiol. 297(3):133-50.
3. Kawamura Y., et al. (1995). Determination of 16RNA sequence of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus Streptococcus. Int J Syst Bacteriol. 45(2):406-8.
4. Bishop C.J., et al. (2009). Assigning strains to bacterial species via the internet. BMC biology.7:3.
5. Jensen A., et al. (2012). Taxonomy of the Anginosus group of the genus Streptococcus and description of *Streptococcus anginosus* subsp. *whileyi* subsp. *nov.* and *Streptococcus constellatus* subsp. *viborgensis* subsp. *nov.* Int. J. Syst. Evol. Microbiol. 63(Pt 7):2506-19.
6. Venkatesan N.K., et al. (2014). Viridans and bovis group streptococci that cause infective endocarditis in two regions with contrasting epidemiology. Int J Med Microbiol. 304(3-4):262-8.
7. Todd E.W., et al. (1928). Variants of hemolytic streptococci; their relation to type-specific substance, virulence, and toxin. J Exp Med. 48(6):751-67.
8. Dagnelie C.F., et al. (1998). Towards a better diagnosis of throat infections (with group A beta-haemolytic streptococcus) in general practice. Br J Gen Pract. 48(427): 959-962.
9. Lancefield R.C. (1933). A serological differentiation of human and other groups of hemolytic streptococci. J. Exp. Med. 57:571.
10. Facklam R. (2002). What Happened to the Streptococci: Overview of Taxonomic and Nomenclature Changes. Clin Microbiol Rev. 15(4):613-30.

11. Shahinas D. (2011). Whole-genome sequence of *Streptococcus pseudopneumoniae* isolate *IS7493*. J. Bacteriol. vol. 193 no. 21 6102-6103.

12. Ferretti J.J., et al. (2001). Complete genome sequence of an M1 strain of *Streptococcus pyogenes* Proc Natl Acad Sci. U S A. 10;98(8):4658-63.

13. Whitworth J.M. (1990). Lancefield group F and related streptococci. J Med Microbiol. 33(3):135-51.

14. Junckerstorffa R.K., et al. (2014). Invasive Streptococcus anginosus group infection does the species predict the outcome? Int J Infect Dis. 18:38-40.

15. Kojima A., et al. (2012). Infection of specific strains of Streptococcus mutans, oral bacteria, confers a risk of ulcerative colitis Sci Rep. 2:332.

16. Gautam M., et al. (2007). *Streptococcus salivarius* bacteremia and spontaneous bacterial peritonitis in liver transplantation candidates. Liver Transpl. 13(11):1582-8.

17. Shelburne S.A., et al. (2014). Streptococcus mitis Strains Causing Severe Clinical Disease in Cancer Patients. Emerg Infect Dis. 20(5):762-71.

18. Sass D.A., et al. (2006). *Streptococcus salivarius* bacteremia in a cirrhotic patient with neutropenia postesophageal variceal ligation. J Clin Gastroenterol. 40(7):654-5.

19. Lonks J.R., et al. (1999). Endocarditis Due to *Streptococcus mitis* with High-Level Resistance to Penicillin and Cefotaxime N Engl J Med. 341:1239.

20. Yew H.S., et al. (2012). Global trends in infective endocarditis epidemiology. Current infectious disease reports. 14, 367-372.

21. Reynolds J.G. (1983). Association of *Streptococcus bovis* bacteremia with bowel disease. Journal of Clinical Microbiology, vol. 17, no. 4, pp. 696-697.

22. Shanan S. (2011). Significant association of *Streptococcus bovis* with malignant gastrointestinal diseases. International Journal of Microbiology Article ID 792019, 5 pages.

23. Klein R.S. (1979). "*Streptococcus bovis* septicemia and carcinoma of the colon," Annals of Internal Medicine. vol. 91, no. 4, pp. 560-562.

24. Boleij A., et al. (2009). Association between *Streptococcus bovis* and colon cancer. Journal of Clinical Microbiology. p. 516 Vol. 47.

25. Stupak A., et al. (2010). The colonization of women genital tract by *Streptococcus agalactiae*. Archives of Perinatal Medicine. 16(1), 48-50.

26. Carapetis J.R., et al. (2005). The global burden of group A streptococcal diseases. Lancet Infect Dis. 5(11):685-94.

27. Juhn Y.J., et al (2012). *Streptococcus pyogenes* upper respiratory infection and atopic conditions other than asthma: a retrospective cohort study. Prim Care Respir J. 21(2):153-8.

28. Cunningham M.W. (2000). Pathogenesis of group A streptococcal infections. Clin Microbiol Rev 13(3):470-511.

29. Stevens D.L. (2000). Streptococcal toxic shock syndrome associated with necrotizing fasciitis. Annu Rev Med. 51:271-88.

30. Schlievert P.M. (1981). Staphylococcal scarlet fever: role of pyrogenic exotoxins. Infect Immun. 31(2): 732-736.

31. Dillon H.C. (1968). Impetigo contagiosa: suppurative and nonsuppurative complication. Clinical, bacteriologic and epidemiologic characteristics of impetigo. Am J Dis Child. 115:530-41.

32. Wannamaker L.W., et al. (1951). Prophylaxis of acute rheumatic fever by treatment of the preceding streptococcal infection with various amounts of depot penicillin. Am J Med. 10:673-95.

33. Falagas M.E., et al. (2008). Effectiveness and safety of short-course vs long-course antibiotic therapy for group a beta hemolytic streptococcal tonsillopharyngitis: a meta-analysis of randomized trials. Mayo Clin Proc. 83 (8): 880-9.

34. Bassetti M., et al. (2000). Erythromycin resistance in *Streptococcus pyogenes* in Italy. Emerg Infect Dis. 6(2): 180-183.

35. Cha S., et al. (2001). The emergence of erythromycin-resistant *Streptococcus pyogenes* in Seoul, Korea. J Infect Chemother. 7(2):81-6.

36. Efstratiou A. (1997). Pyogenic streptococci of Lancefield groups C and G as pathogens in man. Soc Appl Bacteriol Symp Ser. 26:72S-79S.

37. Liu L.C., et al. (2006). Rapid Differentiation between Members of the Anginosus Group and *Streptococcus dysgalactiae* subsp. *equisimilis* within beta-hemolytic Group C and G Streptococci by PCR. J Clin Microbiol. 44(5):1836-8.

38. Reißmann S., et al. (2010). Contribution of *Streptococcus anginosus* to Infections Caused by Groups C and G Streptococci, Southern India. Emerg Infect Dis. 16(4):656-63.

39. Nicholson M.L., et al. (2000). Analysis of immunoreactivity to a *Streptococcus equi* subsp. *zooepidemicus* M-like protein to confirm an outbreak of poststreptococcal glomerulonephritis, and sequences of M-like proteins from isolates obtained from different host species. J Clin Microbiol 38(11):4126-30.

40. Davies M.R., et al. (2007). Phage 3396 from a *Streptococcus dysgalactiae* subsp. *equisimilis* pathovar may have its origins in streptococcus pyogenes. J Bacteriol. Apr;189(7):2646-52.

41. Chen C. and Cleary P.P. (1990). Complete nucleotide sequence ofthe streptococcal C5a peptidase gene of *Streptococcus pyogenes*. Journal of Biological Chemistry. 265, 3161-3167.

42. Cleary P.P., et al. (1991). Virulent human strains of group G streptococci express a C5a peptidase enzyme similar to that produced by group A streptococci. *Infect Immun*. 59(7): 2305-2310.

43. Balke E., et al. (1985). Hyaluronidase activity of b-hemolytic streptococci of the Lancefield group C. Zentralblatt fur Bakteriologie und Hygiene A 259, 194-200.

44. Haidan A., et al. (2000). Pharyngeal carriage of group C and group G streptococci and acute rheumatic fever in an Aboriginal population. Lancet. 356:1167-9.

45. Johnston K.H., et al. (1992). Analysis of the streptokinase gene from group C streptococci (*S. equisimilis* and *S. zooepidemicus*) by the polymerase chain reaction and possible relation to post-streptococcal glomerulonephritis. NewPerspectives on Streptococci and Streptococcal Infections. pp. 343-345.

46. Brandt C.M. and Spellerberg B. (2009). Human Infections Due to *Streptococcus dysgalactiae* Subspecies *equisimilis*. Clin Infect Dis. 49 (5): 766-772.

47. Rantala S., et al. (2010). *Streptococcus dysgalactiae* subsp. *Equisimilis* bacteremia, Finland, 1995-2004. Emerg Infect Dis. 16(5):843-6.

48. McCarty M. (1956). Nature of rheumatic fever. Circulation. 14(6):1138-1143.

49. Wannamaker L.W. (1973). The chain that links the heart to the throat. Circulation. 48(1):9-18.

50. McDonald M., et al. (2006). Low rates of streptococcal pharyngitis and high rates of pyoderma in Australian aboriginal communities where acute rheumatic fever is hyperendemic. Clin Infect Dis. 43(6):683-689.

51. Ogunbi O., et al. (1978). An epidemiological study of rheumatic fever and rheumatic heart disease in Lagos. J Epidemiol Community Health. 32(1):68-71.

52. Watanabe S., et al. (2013). Concomitant regulation of host tissue-destroying virulence factors and carbohydrate. J Infect Dis. 208(9):1482-93.

53. Beachey E.H., et al. (1981). Type-specific protective immunity evoked by synthetic peptide of *Streptococcus pyogenes* M protein. Nature. 292:457-459.

54. Jones K.F., et al. (1991). Genetic diversity among the T-protein genes of group A streptococci. Mol Microbiol. 5(12):2947-52.

55. Beall. B., et al. (1997). Survey of *emm* gene sequences and T-antigen types from systemic *Streptococcus pyogenes* infection isolates collected in San Francisco, California; Atlanta, Georgia; and Connecticut in 1994 and 1995. J Clin Microbiol. 35(5): 1231-1235.

56. Beall B., et al. (1998). Streptococcal *emm*-types associated with T-agglutination types and the use of conserved *emm* gene restriction fragment patterns for subtyping group A streptococci. J. Med. Microbiol. 47:893-898.

57. http://www.cdc.gov/streplab/protocol-*emm*-type.html

58. Blandino G., et al. (2011). *Streptococcus pyogenes emm*-types and subtypes of isolates from paediatric asymptomatic carriers and children with pharyngitis. New Microbiologica. 34, 101-104.

59. Reissmann S., et al. (2012). Region specific and worldwide distribution of collagen-binding M proteins with PARF motifs among human pathogenic streptococcal isolates. PLoS One. 7(1):e30122.

60. Meisal R., et al. (2008). Sequence type and *emm* type diversity in *Streptococcus pyogenes* isolates causing invasive disease in Norway between 1988 and 2003. J Clin Microbiol. 46(6): 2102-2105.

61. McIver K.S., et al. (1999). Regulation of mga transcription in the group A streptococcus: specific binding of mga within its own promoter and evidence for a negative regulator. J Bacteriol. 181(17):5373-83.

62. Gardiner D.L., et al. (1998). Group A streptococcal Vir types are M-protein gene (*emm*) sequence type specific. J Clin Microbiol. 36(4):902-7.

63. McGregor K.F., et al. (2004). Multilocus sequence typing of *Streptococcus pyogenes* representing most known *emm* types and distinctions among subpopulation genetic structures. J Bacteriol. 186(13):4285-94.

64. http://www.mlst.net/

65. Sharma A., et al. (2014). Challenges to developing effective streptococcal vaccines to prevent rheumatic fever and rheumatic heart disease. Dovepress. Volume 4 Pages 39-54.

66. Smeesters PR., et al. (2009). Differences among group A streptococcus epidemiological landscapes: consequences for M protein-based vaccines? Expert Rev Vaccines. 8(12):1705-20.

67. Vlaminckx B.J., et al. (2003). Site-specific manifestations of invasive group a streptococcal disease: type distribution and corresponding patterns of virulence determinants. J Clin Microbiol. 41(11):4941-9.

68. Bessen D.E., et al. (1989). Evidence for two distinct classes of streptococcal M protein and their relationship to rheumatic fever. J Exp Med. 169(1):269-83.

69. Bessen, D.E., et al. (1990). Differentiation between two biologically distinct classes of group A streptococci by limited substitutions of amino acids within the shared region of M protein-like molecules. J. Exp.Med. 172:1757-1764.

70. Bessen, D. E., et al. (1995). Serologic evidence for a class I group A streptococcal infection among rheumatic fever patients. J. Infect. Dis. 172:1608-1611.

71. Massell B.F., et al. (1969). Rheumatic fever following streptococcal vaccination. Report of three cases. JAMA. 207(6): 1115-1119.

72. Johnson D.R., et al. (1992). Epidemiologic analysis of group A streptococcal serotypes associated with severe systemic infections, rheumatic fever, or uncomplicated pharyngitis. J. Infect. Dis. 166:374-382.

73. Stevens D.L. (1992). Invasive group A streptococcus infections. Clin Infect Dis. 14:2-13.

74. Stevens D.L., et al. (1989). Reappearance of scarlet fever toxin A among streptococci in the Rocky Mountain West: Severe group A streptococcal infections associated with a toxic shock-like syndrome. N Eng J Med. 321(1):1-7.

75. Veasy L.G., et al. (1987). Resurgence of acute rheumatic fever in the intermountain area of the United States. N Engl J Med.316:421-427.

76. Smeesters P.R., et al. (2010). The streptococcal M protein: a highly versatile molecule. Trends Microbiol. 18(6):275-82.

77. Gaworzewska E., et al. (1988). Changes in the pattern of infection caused by Streptococcus pyogenes. Epidemiol Infect. 100(2):257-69.

78. Kaplan E.L, et al. (2001). Dynamic epidemiology of group A streptococcal serotypes associated with pharyngitis. Lancet. 358(9290):1334-7.

79. Fischetti V.A (1989). Streptococcal M protein: molecular design and biological behavior. Clin Microbiol Rev. (3):285-314.

80. Beachey, E.H., et al. (1978). Repeating covalent structure of streptococcal M protein. Proc. Natl. Acad. Sci. USA 75:3163-3167.

81. Bisno A.L., et al. (1996). M proteins of group C streptococci isolated from patients with acute pharyngitis. J Clin Microbiol. 34:2511-2515.

82. Collins C.M., et al. (1992). Group G streptococcal M protein exhibits structural features analogous to those of class I M protein of group A streptococci. Infect Immun. 60:3689-3696.

83. McIver K.S., et al. (1999). Regulation of mga transcription in the group A streptococcus: specific binding of mga within its own promoter and evidence for a negative regulator. J Bacteriol. 181(17):5373-83.

84. Bessen D.E., et al. (1996). Genetic correlates of throat and skin isolates of group A streptococci. J Infect Dis. 173: 896-900.

85. Bessen D.E., et al. (1997). Two-domain motif for IgG binding activity by group A streptococcal *emm* gene products. Gene 196: 75-82.

86. Courtney H.S., et al. (2009). Relationship between expression of the family of M proteins and lipoteichoic acid to hydrophobicity and biofilm formation in *Streptococcus pyogenes*. PLoS One. 9;4(1):e4166.

87. Vos P., et al.(2011). Bergey's manual of systematic bacteriology: volume 3: The Firmicutes. pag 682.

88. Phillips G.N. Jr., et al. (1981). Streptococcal M protein: a-Helical coiled-coil structure and arrangement on the cell surface. Proc. NatL Acad. Sci. USA. 78, Vol No. 8, pp. 4689-4693.

89. McArthur D. and Walker M.J. (2006). Domains of group A streptococcal M protein that confer resistance to phagocytosis, opsonization and protection: implications for vaccine development. Molecular Microbiology. 59(1), 1-4, 2005.

90. Fischetti V.A., et al. (1988). Structure, function, and genetics of streptococcal M protein. Rev. Infect. Dis. 10 (Suppl 2), pp. S356-359.

91. Fischetti V.A., et al. (1985). Size variation of the M protein in group A streptococci. J. Exp. Med. 161 pp. 1384-1401.

92. Johansson H.M., et al. (2004). Protein FOG-a streptococcal inhibitor of neutrophil function. Microbiology. 150(Pt 12):4211-21.

93. Hong K. et al (1995). Purification and characterization of M3 protein expressed on the surface of group A streptococcal type 3 strain C203. FEMS Immunol Med Microbiol. 12(1):73-82.

94. Carlsson F., et al. (2006). Signal sequence directs localized secretion of bacterial surface proteins. Nature. 442 pp. 943-946.

95. Haanes-Fritz E., et al. (1988). Comparison of the leader sequences of four group A streptococcal M protein genes. Nucleic. Acids Res. 16 pp. 4667-4677.

96. Marraffini, L.A. et al. (2006). Sortases and the art of anchoring proteins to the envelopes of Gram-positive bacteria. Microbiol. Mol. Biol. Rev. 70, 192-221.

97. Bisno, A.L., et al. (2003). Molecular basis of group A streptococcal virulence. Lancet Infect Dis. 3(4): 191-200.

98. McNamara C. et al. (2008). Coiled-coil irregularities and instabilities in group A Streptococcus M1 are required for virulence. Science. 319, 1405-1408

99. Murakami K. et al. (2008). Structural basis for tropomyosin overlap in thin (actin) filaments and the generation of a molecular swivel by troponin-T. Proc. Natl. Acad. Sci. U. S. A. 105, 7200-7205.

100. Miller L., et al. (1988). Antigenic variation among group A streptococcal M proteins. Nucleotide sequence of the serotype 5 M protein gene and its relationship with genes encoding types 6 and 24 M proteins. J. Biol.Chem. 263, 5668-5673

101. Dinkla K., et al. (2007). Identification of a streptococcal octapeptide motif involved in acute rheumatic fever. J Biol Chem. 282(26):18686-18693

102. Barroso V., et al. (2009). Identification of active variants of PARF in human pathogenic group C and group G streptococci leads to an amended description of its consensus motif. Int J Med Microbiol. 299(8):547-553

103. Horstmann R.D., et al. (1992). Role of fibrinogen in complement inhibition by streptococcal M protein. Infect. Immun. 60:5036-5041.

104. Akesson P., et al. (1994). M1 protein and protein H: IgGFc- and albumin-binding streptococcal surface proteins encoded by adjacent genes. Biochem. J. 300 (877-886).

105. Sandin C., et al (2006). Binding of human plasma proteins to *Streptococcus pyogenes* M protein determines the location of opsonic and non-opsonic epitopes. Mol Microbiol 59: 20-30.

106. Nordenfelt P., et al. (2012). Antibody orientation at bacterial surfaces is related to invasive infection. J Exp Med. 17;209(13):2367-81.

107. Navarre W.W., et al. (1999). Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. Microbiol Mol Biol Rev. 63(1):174-229.

108. Nitsche-Schmitz D.P., et al. (2007). Invasion mechanisms of Gram-positive pathogenic cocci. Thromb Haemost. 98(3):488-96.

109. Kotarsky H., et al. (2001). Group A streptococcal phagocytosis resistance is independent of complement factor H and factor H-like protein 1 binding. Mol Microbiol. 41:817-826.

110. Schmidt K.H., et al.(1993). Multiple binding of type 3 streptococcal M protein to human fibrinogen, albumin and fibronectin. FEMS Immunol Med Microbiol. 7(2):135-43.

111. Retnoningrum, D.S., and Cleary, P.P. (1994). M12 protein from *Streptococcus pyogenes* is a receptor for immunoglobulin G3 and human albumin. Infect Immun. 62: 2387-2394.

112. Blackmore T.K., et al. (1998). M protein of the group A streptococcus binds to the seventh short consensus repeat of human complement factor H. Infect Immun. 66(4):1427-31.

113. Dinkla K., et al. (2003). Rheumatic fever associated *Streptococcus pyogenes* isolates aggregate collagen. J Clin Invest 111(12):1905-1912

114. Frick I.M., et al. (2003). Interactions between M proteins of *Streptococcus pyogenes* and glycosaminoglycans promote bacterial adhesion to host cells. Eur J Biochem. 270(10):2303-11.

115. Oehmcke S., et al. (2010). Streptococcal M proteins and their role as virulence determinants. Clin Chim Acta. 411(17-18):1172-80.

116. Frick I.M., et al. (2000). Virulent aggregates of *Streptococcus pyogenes* are generated by homophilic protein-protein interactions. Mol Microbiol. 37:1232-47.

117. Berggard K., et al. (2001). Binding of human C4 BP to the hypervariable region of M protein: a molecular mechanism of phagocytosis resistance in *Streptococcus pyogenes*. Mol Microbiol. 42:539-551

118. Thern A., et al. (1995). Ig-binding surface proteins of *Streptococcus pyogenes* also bind human C4b-binding protein (C4BP), a regulatory component of the complement system. J. Immunol. 154:375.

119. Johnsson E., et al. (1996). A highly variable region in members of the streptococcal M protein family binds the human complement regulator C4BP. J. Immunol. 157:3021.

120. Whitnack E., et al. (1984). Common protective antigens of group A streptococcal M proteins masked by fibrinogen. J Exp Med. 159:1201-1212

121. Ringdahl U., et al. (2000). A role for the fibrinogen-binding regions of streptococcal M proteins in phagocytosis resistance. Mol Microbiol. 37:1318-1326

122. Onat A., et al. (2011). Complement C3 and cleavage products in cardiometabolic risk. Clin Chim Acta. 412 (13-14): 1171.

123. Herwald H., et al. (2004). M Protein, a classical bacterial virulence determinant, forms complexes with fibrinogen that induce vascular leakage. Cell. 116(3):367-79

124. Oehmcke S., et al. (2009). Treatment of invasive streptococcal infection with a peptide derived from human high-molecular weight kininogen. Blood. 114(2):444-51.

125. Andersen B.R., et al. (1980). Activation of human neutrophil metabolism by streptolysin O. J Infect Dis. 141(5):680-5.

126. Påhlman LI., et al. (2006). Streptococcal M Protein: a multipotent and powerful inducer of inflammation. J Immunol. 177(2):1221-8.

127. Påhlman L.I., et al. (2008). Soluble M1 protein of *Streptococcus pyogenes* triggers potent T cell activation. Cell Microbiol. 10(2):404-14.

128. Haanes, E.J., et al. (1992). Architecture of the vir regulons of group A streptococci parallels opacity factor phenotype and M protein class. J. Bacteriol. 174: 4967.

129. Kehoe M.A., et al. (1996). Horizontal gene transfer among group A streptococci: implications for pathogenesis and epidemiology. Trends Microbiol. 4: 413.

130. Hollingshead, S.K., et al. (1993). Structural heterogeneity of the *emm* gene cluster in group A streptococci. Mol. Microbiol. 8:707-17.

131. Podbielski A. (1993). Three different types of organization of the vir regulon in group A streptococci. Mol. Gen. Genet. 237(1-2):287-300.

132. Caparon M.G., et al. (1987). Identification of a gene that regulates expression of M protein, the major virulence determinant of group A streptococci. Proc Natl Acad Sci. USA. 84: 8677-8681.

133. Podbielski A., et al. (1992). Surface protein-CAT reporter fusions demonstrate differential gene expression in the vir regulon of *Streptococcus pyogenes*. Mol Microbiol. 6: 2253-2265.

134. McIver K.S., et al. (2002). Two DNA-binding domains of Mga are required for virulence gene activation in the group A streptococcus. Mol Microbiol. 43: 1591-1601.

135. Thern A., et al. (1998). Expression of two different antiphagocytic M Proteins by *Streptococcus pyogenes* of the OF+ lineage. The Journal of Immunology. vol. 160 no. 2 860-869.

136. Podbielski A., (1996). M-related protein (Mrp) contributes to group A streptococcal resistance to phagocytosis by human granulocytes. Mol Microbiol. (3):429-41.

137. Bessen, D.E., et al. (1992). Nucleotide sequences of two adjacent M or M-like protein genes of group A streptococci: different RNA transcript levels and identification of a unique immunoglobulin A-binding protein. Infect Immun. 60: 124.

138. Jeppson, H., et al. (1992). Duplication of a DNA sequence homologous to genes for immunoglobulin receptors and M proteins in *Streptococcus pyogenes*. FEMS Microbiol. Lett. 71: 139.

139. Podbielski, A., et al. (1995). The group A streptococcal virR49 gene controls expression of four structural vir regulon genes. Infect Immun. 63(1):9-20.

140. Ji Y., et al. (1998). Impact of M49, Mrp, Enn, and C5a peptidase proteins on colonization of the mouse oral mucosa by *Streptococcus pyogenes*. Infect Immun. 66(11):5399-405.

141. James B., et al. (1999). New protective antigen of group A streptococci. J Clin Invest. Volume 103, Issue 9 103(9):1261-1268.

142. Smoot J.C., et al. (2002). Molecular analysis of group A Streptococcus type *emm*18 isolates temporally associated with acute rheumatic fever outbreaks in Salt Lake City, Utah. J Clin Microbiol. 40(5):1805-1810.

143. Ahmed E.A. (2010). Streptococcal protective antigens (Spa): a new family of type-specific proteins of group A streptococci. Eur J Clin Microbiol Infect Dis. 29(1):51-7.

144. Niedermeyer S.E., et al. (2014). Group a streptococcus expresses a trio of surface proteins containing protective epitopes. Clin Vaccine Immunol. (10):1421-5.

145. McLellan D.G. (2001). Spa contributes to the virulence of type 18 group A streptococci. Infect Immun. 69(5):2943-9.

146. Silva F.G. (1998). Acute postinfectious glomerulonephritis and glomerulonephritis complicating persistent bacterial infection. Hepinstall's pathology of the kidney. 5th ed. Lippincott-Raven Publishers, Philadelphia, Pa. p. 389-453.

147. Bisno, A.L. (1970). Contrasting epidemiology of acute rheumatic fever and acute glomerulonephritis: nature of the antecedent streptococcal infection. N. Engl. J. Med. 283:561-565.

148. White, A.V., et al. (2001). Childhood post-streptococcal glomerulonephritis as a risk factor for chronic renal disease in later life. Med J. 174(10): p. 492-6.

149. Streeton, C.L., et al. (1995). An epidemic of acute post-streptococcal glomerulonephritis among aboriginal children. J Paediatr Child Health. 31(3): p. 245-8.

150. Bisno A. L. (1995). Non-suppurative poststreptococcal sequelae: rheumatic fever and glomerulonephritis. Principles and practice of infectious diseases. vol.2. p. 1799-1810.

151. Khandke K.M., et al. (1987). Difference in the structural features of streptococcal M proteins from nephritogenic and rheumatogenic serotypes. J Exp Med. 166 (1):151-62.

152. Goroncy-Bermes P., (1987). Monoclonal antibody to human renal glomeruli cross-reacts with streptococcal M protein. Infect Immun. 55:2416-2419.

153. Lange C.F. (1969). Chemistry of cross-reactive fragments of streptococcal cell membrane and human glomerular basement membrane. Transplant. Proc. 1:959-963.

154. Kraus W. (1990). Identification of an epitope of type 1 streptococcal M protein that is shared with a 43-kDa protein of human myocardium and renal glomeruli. J. Immunol. 145:4089-4093.

155. Christensen P., et al. (1979). Reevaluation of experiments intended to demonstrate immunological cross-reactions between the mammalian tissues and streptococci. Prog Allergy. 26: 1-41. 56.

156. Cochrane C.G. (1971). Mechanisms involved in the deposition of immune complexes in tissues. J Exp Med. 134(3):75-89.

157. Fish A.J., et al. (1966). Acute serum sickness nephritis in the rabbit. An immune deposit disease. Am J Pathol. 49(6):997-1022.

158. Friedman J. (1984). Immunological studies of post-streptococcal sequelae: evidence for presence of streptococcal antigens in circulating immune complexes. J. Clin. Investig. 74:1027-1034.

159. Nitsche-Schmitz D.P., et al. (2012). Host-pathogen interactions in streptococcal immune sequelae. Curr Top Microbiol Immunol. 368:155-71.

160. Chhatwal G.S., et al. (2008). Streptococcal diseases. Encyclopedia of public health. vol 6, Academic Press, San Diego, pp 231-241.

161. Uziel Y., et al. (2011). Post-streptococcal reactive arthritis in children: a distinct entity from acute rheumatic fever. Pediatr Rheumatol Online J. 9(1):32.

162. Carapetis J.R., et al. (2005). Acute rheumatic fever. Lancet. 366(9480):155-68.

163. Jones T. Duckett. (1944). The diagnosis of rheumatic fever. JAMA. 126 (8): 481-4.

164. Virmani R., et al. (1999). Amer. Reg. Path. 217-234. AFIP, Washington DC.

165. Gulizia J.M., et al. (1999). Rheumatic Fever 235-244 (Amer. Reg. Path. AFIP, Washington DC).

166. Tandon R., et al. (2013). Revisiting the pathogenesis of rheumatic fever and carditis. Nature Reviews Cardiology. 10, 171-177.

167. Roberts W.C. and Virmani R., (1978). Aschoff bodies at necropsy in valvular heart disease. Evidence from an analysis of 543 patients over 14 years of age that rheumatic heart disease, at least anatomically, is a disease of mitral valve. Circulation. 57, 803-807.

168. Kaplan M.H., et al. (1964). Presence of bound immunoglobulins and complement in the myocardium in acute rheumatic fever. Association with cardiac failure. N Engl J Med 271:637-645.

169. Cunningham M.W. (2003). Autoimmunity and molecular mimicry in the pathogenesis of poststreptococcal heart disease. Front Biosci. 8:533-543.

170. Guilherme L., et al. (2006). Molecular mimicry in the autoimmune pathogenesis of rheumatic heart disease. Autoimmunity. 39(1):31-39.

171. Roberts S., et al. (2001). Pathogenic mechanisms in rheumatic carditis: focus on valvular endothelium. J Infect Dis. 183(3):507-511.

172. Goldstein I., et al. (1968). Isolation from heart valves of glycopeptides which share immunological properties with Streptococcus haemolyticus group A polysaccharides. Nature. 219(5156):866-868.

173. Wannamaker L.W. (1973). The chain that links the heart to the throat. Circulation 48(1):9-18.

174. Cunningham M.W. (2012). Streptococcus and rheumatic fever. Curr Opin Rheumatol 24(4):408-416.

175. Eriksson B.K., et al. (1999). Invasive group A streptococcal infections: T1M1 isolates expressing pyrogenic exotoxins A and B in combination with selective lack of toxin-neutralizing antibodies are associated with increased risk of streptococcal toxic shock syndrome. J Infect Dis. 180(2):410-418.

176. Kotb M. (1995). Bacterial pyrogenic exotoxins as superantigens. Clin Microbiol Rev. 8(3):411-426.

177. Sriskandan S., et al. (1996). Detection of circulating bacterial superantigen and lymphotoxin-alpha in patients with streptococcal toxic-shock syndrome. Lancet 348(9037):1315-1316.

178. Proft T., et al. (2003). Bacterial superantigens. Clin Exp Immunol. 133(3):299-306.

179. Kaplan M.H., et al. (1964). Immunologic relation of streptococcal and tissue antigens. ii. Cross-reaction of antisera to mammalian heart tissue with a cell wall constituent of certain strains of group a streptococci. J Exp Med. 119:643-650.

180. Zabriskie J.B., et al. (1966). An immunological relationship between the group. A streptococcus and mammalian muscle. J Exp Med. 124(4):661-678.

181. Dudding B.A., et al. (1968). Persistence of streptococcal group A antibody in patients with rheumatic valvular disease. J Exp Med. 128(5):1081-1098.

182. Veasy L.G., et al. (2004). Temporal association of the appearance of mucoid strains of *Streptococcus pyogenes* with a continuing high incidence of rheumatic fever in Utah. Pediatrics. 113(3 Pt 1):168-172.

183. Blank M., et al. (2007). Molecular mimicry and auto-immunity. Clin Rev Allergy Immunol. 32(1):111-118.

184. Guilherme L., et al. (2010). Rheumatic fever and rheumatic heart disease: cellular mechanisms leading autoimmune reactivity and disease. J Clin Immunol. 30(1):17-23.

185. McCormack J.M., et al. (1993). Poststreptococcal anti-myosin antibody idiotype associated with systemic lupus erythematosus and Sjogren's syndrome. J Infect Dis. 168(4):915-921.

186. Zabriskie J.B. (1967). Mimetic relationships between group A streptococci and mammalian tissues. Adv Immunol. 7:147-188.

187. Bronze M.S., et al. (1988). Protective and heart-crossreactive epitopes located within the NH2 terminus of type 19 streptococcal M protein. J Exp Med 167(6):1849-1859.

188. Cunningham M.W., et al. (1989). Human and murine antibodies cross-reactive with streptococcal M protein and myosin recognize the sequence GLN-LYS-SER-LYS-GLN in M protein. J Immunol. 143(8):2677-2683.

189. Faé KC., et al. (2005). How an autoimmune reaction triggered by molecular mimicry between streptococcal M protein and cardiac tissue proteins leads to heart lesions in rheumatic heart disease. J Autoimmun. 24(2):101-9.

190. Stollerman G.H. (1969). Nephritogenic and rheumatogenic group A streptococci. J Infect Dis. 120(2):258-263.

191. Chan A.L., et al. (2011). Cutting edge issues in Goodpasture's disease. Clin Rev Allergy Immunol. 41(2):151-162.

192. Rodriguez-Iturbe B., et al. (2007). Pathogenesis of poststreptococcal glomerulonephritis a century after Clemens von Pirquet. Kidney Int. 71(11):1094-1104.

193. Steer A.C., et al. (2009). Group A streptococcal vaccines: facts versus fantasy. Curr Opin Infect Dis. 22(6):544-552.

194. Steer AC., et al. (2009). Global *emm* type distribution of group A streptococci: systematic review and implications for vaccine development. Lancet Infect Dis. 9(10):611-616.

195. Kreikemeyer B., et al. (2005). *Streptococcus pyogenes* Collagen Type I-binding Cpa Surface Protein: expression profile, binding characteristics, biological functions, and potential clinical impact. J. Biol. Chem. 280:33228-33239.

196. Bober M., et al. (2011). The membrane bound LRR lipoprotein Slr, and the cell wall-anchored M1 protein from *Streptococcus pyogenes* both interact with type I collagen. PLoS One. 6(5):e20345.

197. Broyles L.N., et al. (2009). Population-based study of invasive disease due to beta-hemolytic streptococci of groups other than A and B. Clin Infect Dis. 48(6):706-712.

198. Reißmann S., et al. (2010). Contribution of *Streptococcus anginosus* to infections caused by groups C and G streptococci, southern India. Emerg Infect Dis. 16(4):656-663.

199. Bramhachari P.V., et al. (2010). Disease burden due to *Streptococcus dysgalactiae* subsp. *equisimilis* (group G and C streptococcus) is higher than that due to *Streptococcus pyogenes* among Mumbai school children. J. Med Microbiol. 59(Pt 2):220-223.

200. Dale J.B., et al. (2013). Group A streptococcal vaccines: paving a path for accelerated development. Vaccine. 31(Suppl 2):B216-B222.

201. Widdowson, J.P., et al. (1976). An M-associated protein antigen (MAP) of group A streptococci. J. Hyg. 69:553-564.

202. Wilson, M.G., et al. (1943). The familial epidemiology of rheumatic fever. J. Pediatr. 22:468-492.

203. Henningham A., et al. (2013). Group a streptococcal vaccine candidates: potential for the development of a human vaccine. Curr Top Microbiol Immunol. 368:207-42.

204. O'Connor S.P., et al. (1987). In vivo *Streptococcus pyogenes* C5a peptidase activity: analysis using transposon- and nitrosoguanidine-induced mutants. J Infect Dis. 156(3):495-504.

205. Wexler D.E., et al. (1985). Mechanism of action of the group A streptococcal C5a inactivator. Proc Natl Acad Sci. U S A. 82(23):8144-8148.

206. Husmann L.K., et al. (1997). Role of putative virulence factors of *Streptococcus pyogenes* in mouse models of long-term throat colonization and pneumonia. Infect Immun. 65(4):1422-30.

207. Chmouryguina I., et al. (1996). Conservation of the C5a peptidase genes in group A and B streptococci. Infect Immun. 64(7):2387-2390.

208. Cleary P.P., et al.(2004). Immunization with C5a peptidase from either group A or B streptococci enhances clearance of group A streptococci from intranasally infected mice. Vaccine. 22(31-32):4332-4341.

209. Koroleva I.V., et al. (2002). Structural Heterogeneity of the Streptococcal C5a Peptidase Gene in *Streptococcus pyogenes.* J Bacteriol. 184(22): 6384-6386.

210. Shet A., et al. (2003). Immune response to group A streptococcal C5a peptidase in children: implications for vaccine development. J Infect Dis. 188(6):809-817.

211. Roggiani M., et al. (2000). Toxoids of streptococcal pyrogenic exotoxin A are protective in rabbit models of streptococcal toxic shock syndrome. Infect Immun. 68(9):5011-5017.

212. McCormick J.K., et al. (2000). Development of streptococcal pyrogenic exotoxin C vaccine toxoids that are protective in the rabbit model of toxic shock syndrome. J Immunol. 165(4): 2306-2312.

213. Bhatnagar A., et al. (1999). Superantigen-induced T cell responses in acute rheumatic fever and chronic rheumatic heart disease patients. Clin Exp Immunol. 116(1):100-6.

214. Brocke S., et al. (1993). Induction of relapsing paralysis in experimental autoimmune encephalomyelitis by bacterial superantigen. Nature. 1993;365(6447):642-644.

215. Cole B.C., et al. (1993). Triggering and exacerbation of autoimmune arthritis by the Mycoplasma arthritidis superantigen MAM. Arthritis Rheum. 36(7):994-1002.

216. Fox E.N., et al. (1973). Protective study with a group A streptococcal M protein vaccine. Infectivity challenge of human volunteers. J Clin Invest. 52(8): 1885-1892.

217. Dale J.B., et al. (1985). Multiple, heart-cross-reactive epitopes of streptococcal M proteins. J Exp Med 161(1):113-122.

218. Cunningham M.W., et al. (1997). Molecular analysis of human cardiac myosin-cross-reactive Band T-cell epitopes of the group A streptococcal M5 protein. Infect Immun. 65:3913-3923.

219. Hu M.C., et al. (2002). Immunogenicity of a 26-valent group A streptococcal vaccine. Infect Immun. 70(4):2171-2177.

220. McNeil S.A., et al. (2005). Safety and immunogenicity of 26-valent group A streptococcus vaccine in healthy adult volunteers. Clin Infect Dis. 2005;41(8):1114-1122.

221. Dale J.B., et al. (2011). New 30-valent M protein-based vaccine evokes cross-opsonic antibodies against non-vaccine serotypes of group A streptococci. Vaccine. 29(46):8175-8178.

222. Dey N., et al. (2005). High diversity of group A Streptococcal *emm* types in an Indian community: the need to tailor multivalent vaccines. Clin Infect Dis. 40(1):46-51.

223. Bessen D., et al. (1988). Influence of intranasal immunization with synthetic peptides corresponding to conserved epitopes of M protein on mucosal colonization by group A streptococci. Infect Immun. 56(10):2666-2672.

224. Bessen D., et al. (1990). Synthetic peptide vaccine against mucosal colonization by group A streptococci. i. Protection against a heterologous M serotype with shared C repeat region epitopes. J Immunol. 145(4):1251-1256.

225. Guilherme L., et al. (2006). Towards a vaccine against rheumatic fever. Clin Dev Immunol. 13(2-4):125-132.

226. Guilherme L., et al. (2009). A vaccine against *S. pyogenes*: design and experimental immune response. Methods. 2009;49(4):316-321.

227. Postol E., et al. (2013). StreptInCor: a candidate vaccine epitope against *S. pyogenes* infections induces protection in outbred mice. PLoS One. 8(4):e60969.

228. Guerino MT., et al. (2011). HLA class II transgenic mice develop a safe and long lasting immune response against StreptInCor, an anti-group A streptococcus vaccine candidate. Vaccine. 29(46):8250-8256.

229. Brandt E.R., et al. (1997). Human antibodies to the conserved region of the M protein: opsonization of heterologous strains of group A streptococci. Vaccine. 15(16): 1805-1812.

230. Moyle P.M., et al. (2013). An efficient, chemically-defined semisynthetic lipid-adjuvanted nanoparticulate vaccine development system. Nanomedicine. 9(7):935-44.

231. Zhong W., et al. (2009). Development of conformational mimetics of conserved *Streptococcus pyogenes* minimal epitope as vaccine candidates. Curr Drug Deliv. 6(5):520-7.

232. Brandt E.R., et al. (1996). Opsonic human antibodies from an endemic population specific for a conserved epitope on the M protein of group A streptococci. Immunology. 89(3):331-337.

233. Penfound T.A., et al. (2010). Protective efficacy of group A streptococcal vaccines containing type-specific and conserved M protein epitopes. Vaccine. 28(31):5017-22.

234. Bauer M.J., et al. (2012). Evaluation of novel *Streptococcus pyogenes* vaccine candidates incorporating multiple conserved sequences from the C-repeat region of the M-protein. Vaccine. 30(12):2197-2205.

235. Sharma A., et al. (2013). Identification of potential universal vaccine candidates against group A Streptococcus by using high throughput in silico and proteomics approach. J Proteome Res. 12(1):336-346.

236. Rappuoli R., et al. (2000). Reverse vaccinology. Curr Opin Microbiol. 3(5): 445-450.

237. Chunget C. T., et al. (1989). One-step preparation of competent *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. Proc. Natl. Acad. Sci. USA. Vol. 86, pp. 2172-2175.

238. pQE-30TEV figure was adapted from http://www.qiagen.com

239. Hunter, W., et al. (1962). Preparation of iodine-131 labelled human growth hormone of high specificity. Nature. 194:495-506.

240. Abdiche Y., et al. (2008). Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet. Anal Biochem. 377(2):209-17.

241. Dayne D. (2012). BioLayer Interferometry (BLI). How Does it Work?. Spring, volume 5-issue 1.

242. Hitzmann A. (2009). Untersuchungen an pathogenitätsassoziierten Proteininteraktionen des M-Proteins FOG aus *Streptococcus dysgalactiae equisimilis.* Master Thesis. TU Braunschweig.

243. Guidelines for the diagnosis of rheumatic fever. Jones Criteria, 1992 update. Special writing group of the committee on rheumatic fever, endocarditis, and kawasaki disease of the council on cardiovascular disease in the young of the american heart association. Jama. 268(15):2069-73.

244. Frey A., et al. (1998). A statistically defined endpoint titer determination method for immunoassays. J Immunol Methods. 221 35-41.

245. Courtney H.S., et al. (2003). Serum opacity factor (SOF) of *Streptococcus pyogenes* evokes antibodies that opsonize homologous and heterologous SOF-positive serotypes of group A streptococci. Infect Immun. 71(9):5097-103.

246. Uchiyama S., et al. (2013). Coiled-coil irregularities of the M1 protein structure promote M1-fibrinogen interaction and influence group A Streptococcus host cell interactions and virulence. J Mol Med. 91(7):861-9.

247. Dinkla K., et al. (2009). Crucial role of the CB3-region of collagen IV in PARF-induced acute rheumatic fever. PLoS ONE. 4(3): e4666.

248. Nitsche-Schmitz D.P., et al. (2007). Group G streptococcal IgG binding molecules FOG and protein G have different impacts on opsonization by C1q. J Biol Chem. 282(24): 17530-17536.

249. Gupta, M., et al. (2002). Serum cardiac troponin I in acute rheumatic fever. Am J Cardiol. 89(6): 779-782.

250. Galvin J.E., et al. (2000). Cytotoxic mAb from rheumatic carditis recognizes heart valves and laminin. J Clin Invest. 106(2): 217-224.

251. Yurchenco, P.D. and Schittny J.C. (1990). Molecular architecture of basement membranes. Faseb J. 4(6): 1577-1590.

252. McDonnell A.V., et al. (2006). Paircoil2: Improved prediction of coiled coils from sequence. Bioinformatics. Vol. 22(3).

ACKNOWLEDGMENTS

I would like to thank to my PhD supervisor, Dr Daniel Patric Nitsche-Schmitz for supporting me during the past three years. Patric has been a great mentor, always keen on to give me suggestions and scientific advices. He has been supportive and has given me the freedom to pursue various projects and establish good collaboration during my PhD.

I am also very thankful to thank Prof. Gurusharan Singh Chhatwal, a pioneer in streptococcal research, who has welcomed me into his research group and has challenged me to do my best as a scientist.

I also would like to thank the members of my PhD committee, Dr Daniel Patric Nitsche-Schmitz, Prof. Michael Steinert, Prof. Manfred Rohde and Prof. Gurusharan Singh Chhatwal for their time, enthusiasm, suggestions and advices to improve my PhD project.

A special thank goes to my friend and colleague Venkatesan Naveen Kumar. Naveen has collaborated with me in several exciting projects during the last year and I am also thankful to the several scientific discussions we had together about streptococci and new research opportunities in this field. I would like to thank Dr. René Bergmann for being always available and keen on to help me. He has collaborated with me in many projects and always gave me good wet lab suggestions. I am also very grateful to Dr. Satoshi Uchiyama for sharing his knowledge and ideas through fruitfull discussions and collaborative research.

Astonishing electron microscopy images of chain forming streptococci would not be possible without Prof. Manfred Rohde. Therefore, I am very grateful for his help and great career advices.

A special thanks goes to Katja Mummenbrauer, a lovely colleague that has been always helpful and available since the first day I arrived in the lab D1.35.

I am also very thankful to the HZI Graduate School for the financial support and for the several career development opportunities and soft skills courses provided.

My gratitude to all members of the MMIK and MINP groups for being a fantastic “piece of Germany” that I will always bring along with me.

I want to thank the TT (Thirsty Thursday) people because you have been my family in Germany. In fact, I would need a full book to describe how important you all have been in these last three years (basically, you helped me to survive through my PhD).

Infinite thankfulness goes to my muse and my love, Márcia. You have helped and motivated me like no other. You are the best person I ever meet and I cannot imagine a single day of this 3-year journey without you.

Finally, I would like to thank my parents for all the support provided over all these years and for motivating me to always follow my passions. An endless appreciation to my grandparents, who made me fell at home despite the big distance. A very special thanks goes to my brother and sister as they are the ones that can cheer me up at any occasion.

SUPPLEMENTS

<u>Supplement A:</u> FASTA files of the recombinant proteins used in this work. Molecular weight is expressed in kDa and was calculated using ProtParam. http://web.expasy.org/cgi-bin/protparam/protparam

>StG11 (Fog) (65.247 kDa)

HHHHHHGENLYFQGSAENTYDRWKAQTEEARTDKLIAGFANLDADVTNLGKMMDELQKLKDFSKQNNSIGEYAR
YLQKLNDQFQEYYEQVVGDDSRRVLAKELAKNTELNEKLSELSKTSQALAKELQEQKENYDLVKTVHADTVKKHQ
KLVDEITKKLGEEETERHSLQEELNKAQQELAQKQQELKDKQADYDLVVETHAYTVKEHQKLVDEITKKLGEEETE
RHSLQEELNKAQQELAQKQQELEAEKLAKEGIVDSLTAYVTEKEAEVKKLTDSLAAKDAEIQEKEAEKDRQQHMY
EAFMSQYKEKVEKQEQELAKLKQLETINNNLLGNAKDMIAKLSAKNEQLASDKAKLEEQNKISDASRKGLRRDLNA
SREAKKQLEAEHQKLEEQNKISEASRKGLRRDLDASREAKKQVEKDLANLTAELDKVKEDKQISDASRKGLRRDL
DASREAKKQVEKALEEANSKLAALEKLNKELEESKKLTEKEKAELQAKLEAEAKALKEKLAKQAEELAKLRAGKAS
DSQTPEATPGNKVVPGKGQASQAGTKPNQNKEPMKETKRQLPSTG

>StG11 AA (Fog AA) (65.247 kDa)

HHHHHHGENLYFQGSAENTYDRWKAQTEEARTDKLIAGFANLDADVTNLGKMMDELQKLKDFSKQNNSIGEYAR
AAQKLNDQFQEYYEQVVGDDSRRVLAKELAKNTELNEKLSELSKTSQALAKELQEQKENYDLVKTVHADTVKKH
QKLVDEITKKLGEEETERHSLQEELNKAQQELAQKQQELKDKQADYDLVVETHAYTVKEHQKLVDEITKKLGEEET
ERHSLQEELNKAQQELAQKQQELEAEKLAKEGIVDSLTAYVTEKEAEVKKLTDSLAAKDAEIQEKEAEKDRQQHM
YEAFMSQYKEKVEKQEQELAKLKQLETINNNLLGNAKDMIAKLSAKNEQLASDKAKLEEQNKISDASRKGLRRDLN
ASREAKKQLEAEHQKLEEQNKISEASRKGLRRDLDASREAKKQVEKDLANLTAELDKVKEDKQISDASRKGLRRD
LDASREAKKQVEKALEEANSKLAALEKLNKELEESKKLTEKEKAELQAKLEAEAKALKEKLAKQAEELAKLRAGKA
SDSQTPEATPGNKVVPGKGQASQAGTKPNQNKEPMKETKRQLPSTG

>StG11 1A (Fog1-A) (15.533 kDa)

AENTYDRWKAQTEEARTDKLIAGFANLDADVTNLGKMMDELQKLKDFSKQNNSIGEYARYLQKLNDQFQEYYEQ
VVGDDSRRVLAKELAKNTELNEKLSELSKTSQALAKELQEQKENYDLVKTVHADTVKKHQ

>M1 (AP1) (48.608 kDa)

HHHHHHGENLYFQGSNGDGNPREVIEDLAANNPAIQNIRLRHENKDLKARLENAMEVAGRDFKRAEELEKAKQA
LEDQRKDLETKLKELQQDYDLAKESTSWDRQRLEKELEEKKEALELAIDQASRDYHRATALEKELEEKKKALELAI
DQASQDYNRANVLEKELETITREQEINRNLLGNAKLELDQLSSEKEQLTIEKAKLEEEKQISDASRQSLRRDLDASR
EAKKQVEKDLANLTAELDKVKEDKQISDASRQGLRRDLDASREAKKQVEKDLANLTAELDKVKEEKQISDASRQG
LRRDLDASREAKKQVEKALEEANSKLAALEKLNKELEESKKLTEKEKAELQAKLEAEAKALKEQLAKQAEELAKLR
AGKASDSQTPDTKPGNKAVPGKGQAPQAGTKPNQNKAPMKETKRQLPSTG

>M55 (56.463 kDa)

HHHHHHGENLYFQGSQTEPSQTNNRLYQERQRLQDLKSKFQDLKNRSEGYIQQYYDEEKNSGSNSNWYATYLK
ELNDEFEQAYNELSGDGVKKLAASLMEERVALRDEIDQIKKISEELKNKLRAKEEELKNKKEERELEHAAYAADAK
KHEEYVKSMSLVLMDKEEERHKLEQSLDTAKAELVKKEQELQLVKGNLDQKEKELENEELAKESAISDLTEQITAK
KAEVEKLTQDLAAKSAEIQEKEAEKDRQQHMYEAFMSQYKEKVEKQEQELAKLKQLETINNNLLGNAKDMIAKLS
AENEQLASDKAKLEEQNKISEASRKGLRRDLDASREAKKQVEKDLANLTAELDKVKEDKQISDASRKGLRRDLDA
SREAKKQVEKSLEEANSKLAALEKLNKELEESKKLTEKEKAELQAKLEAEAKALKEQLAKQAEELAKLRAGKASDS
QTPDAKPGNKVVPGTGQAPQAGTKPNQNKAPMKETKRQLPSTG

>StG2078 (63.128 kDa)

HHHHHHGENLYFQGSVENTAQYDRQKAETEKNRIYGLERNLLDLKAQVDHLENLMDNLQTQKAFVNNLSWSDN
NIGEYAKLLEELNDQFEEYYEKVVGDDSRRVFAKVIEENKGLNEKLSGLLEKSNDLAKKLKEKEEELKNKEAERDL
EHAAYARDRKEHEEYVRRVGLLLEDRAKEKQELQSSVEQAKKKIDSLNQELAQKEEELAKEKLAKQGIIDSTSDYV
TEKESKIAKLEEEIQTNKAELESLNSQLESLNSNLDKKTSEIDNLKQELTRKENMYESFLNQAKENLASKEQELXKL
KQLETINNNLLGNAKDMIAKLSAKSEQLASDRAKLEEQNKISDASRQSLRRDLNASREAKKQLEAEHQKLEEQNKI
SEASRQGLRRDLNASREAKKQVEKDLANLTAELDKVKEEKQISDASRQGLRRDLEASREAKKQVEKALEEANSKL
AALEKLNKELEESKKLTEKEKAELQAKLEAEAKALREKLAKQAEELAKLRAGKASDSQTPDATPGNKVVPGKGQA
PQAGTKPNQNKAPMKETKRQLPSTG

>M3.23 (59434 kDa)

HHHHHHGENLYFQGSDARSVNGEFPRHVKLKNEIENLLDQVTQLYNKHNSNYQQYNAQAGRLDLRQKAEYLKG
LNDWAERLLQELNGEDVKKVLGKVAFEKDDLEKEVKELKKKIDKKEKEYQDLDKDFDLAKQGHVLSDKRHQQELE
EKEKKVTEATAKVGQISEELETVKQKVESTMQDLTEKQNRVSQLEQELATTKQNAKEDFELAALANAADKQKLEA
KIADLETKLKEAKEDFELAALGHQHAHNEYQAKLAEKDDQIKQLEEQKQILDASRKGTARDLEAVRQAKKATEAEL
NNLKAELAKVTEQKQILDASRKGTARDLEAVRQAKAQVEAALKQLEEQNRISEASRKGLRRDLDASREAKKQVEK
DLANLTAELDKVKEEKQISDASRQGLRRDLDASREAKKQVEKALEEANSKLAALEKLNKELEESKKLTEKEKAELQ
AKLEAEAKALKEQLAKQAEELAKLRAGKASDSQIPDTKPGNKAVPGKGQAPQAGTKPNQNKAPMKETKRQLPST
G

>M3.0 (59.448 kDa)

HHHHHHGENLYFQGSDARSVNGEFPRHVKLKNEIENLLDQVTQLYTKHNSNYQQYNAQAGRLDLRQKAEYLKGL
NDWAERLLQELNGEDVKKVLGKVAFEKDDLEKEVKELKEKIDKKEKEYQDLDKDFDLAKQGYVLSDKRHQQELE
EKEKKVTEATAKVGQISEELETVKQKVESTMQDLTEKQNRVSQLEQELATTKQNAKEDFELAALANAADKQKLEA
KIADLETKLKEAKEDFELAALGHQHAHNEYQAKLAEKDDQIKQLEEQKQILDASRKGTARDLEAVRQAKKATEAEL
NNLKAELAKVTEQKQILDASRKGTARDLEAVRQAKAQVEAALKQLEEQNRISEASRKGLRRDLDASREAKKQVEK
DLANLTAELDKVKEEKQISDASRQGLRRDLDASREAKKQVEKALEEANSKLAALEKLNKELEESKKLTEKEKAELQ
AKLEAEAKALKEQLAKQAEELAKLRAGKASDSQIPDTKPGNKAVPGKGQAPQAGTKPNQNKAPMKETKRQLPST
G

>StG120.1s (16.660 kDa)

HHHHHHGENLYFQGSEENITQYDKWNWETQFGRTSDLTSKLQLLKDRVSGLEYIMNHLNSEIRPDKWQSHDLGV
LANYLKTLNDHFEENYEKVVGDDSRRVLAQEVEENRALKEKLSELTTKSEKLANELKAQQEIYQLAV

>PARF tandem (38.037 kDa)

HHHHHHGENLYFQGSLQQAEYLQRLNDHFEDVYTEYLKRLNDDAEAGRTEYLKSLNDWAEEQKAEYLRSLNDW
AEDTHTQYLKRLNDYFQEVYTQYLKRLNDQFQASVAVALTVLGAGLAKTYADYLKTLNDRFENTYTEYLKSLNDH
FQKDYANYLKELNDHFENWYATYLKELNDEFERQKAEYLKGLNDWAETQKAEYLKNLNDWAEGQMAEYLKALN
DHFEGVLANYLKTLNDHFEGEYARYLQKLNDQFQGVYARYLKKLNDQFQGVYARYLETLNDRFQGVYARYLKRL
NDHFQSDDEAYLKRLNEWAE

>Spa18 (57704 kDa)

HHHHHHGENLYFQGSDSVSGLEVADPSDSKKLIELGLAKYLNDKLPFKTKEDSEILSELRDVLKNANPETLKSLLN
GMDQGHISFSDRNNRYNRLSQYINSFRKDDDDYLHNGYSLGSLVIEAIKYRLDSESHLKEELLKQTAELEQRKNAE
VDLKSEKKRLEAQIEKVGYDIANKQQELEKARSDQKELSESIQKLTSRFKKESDAKQKELDEAKAANKSLSESATK
TLARSSKITNELKDKLAASEKDKNRAFQVSSELANKLHETETSRDKALAESKELADKLAVKTAEAEKLMENVGSLD
RLVESAKREMAQKLAEIDQLTADKAKADAELAAANDTIASLQTELEKVKTELAVSERLIESGKREIAELEKQKDASD
KALAESQANVAELEKQKAASDAKVAELEKEVEAAKAEVADLKAQLAKKEEELEAVKKEKEALEAKIEELKKAHAEE
LSKLKEMLEKKDHANADLQAEINRLKQELADRIKSLSQGGRASQTNPGSTTAKAGQLPSTG

>Spa36 (55.381 kDa)

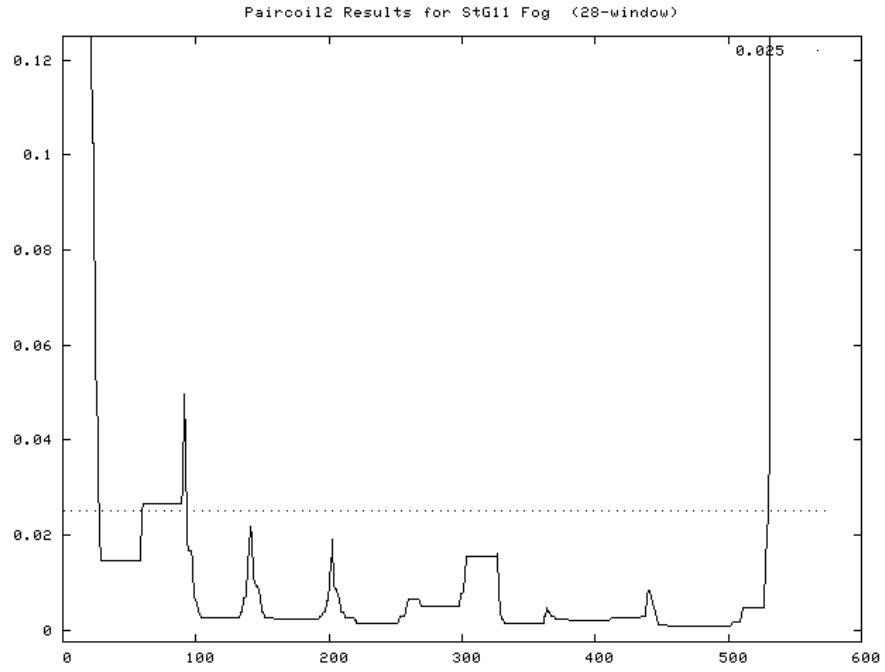
HHHHHHGENLYFQGSEDRINSDINRRIVDKKDAENLRNAIIAKGARLKNDSEVISFLREALSNATVDTLATLLSGIDP
ARFGYSNSQLEFYSRQLGSLNSGVSDWQQGSVNLKTLLIEEFAKRIKSEQKLKEVLAEQAAELELRKSEESELKTQ
KERLESKLENAEYATAIKQKELDDAKEANKTLSESIAKTLSRSTKEKDKLKEELVKEKTKAAKIAKELMDKLTASEKD
KDRAIQITTELTNRLHETEASRDKAFAVSTELANKLAAKTAEAEKLMQNVGSLDRLVESAKREIAEKLAEIDQLTAD
KAKADAELATANDTIASLQTELEKAKTELAVSERLIESGKREMAELQKQKDASDKALAESQANVAELEKQKAASDA
KVAELEKEVEAAKAEVADLKAQLAKKEEELEAVKKEKEALEAKIEELKKAHAEELSKLKEMLEKKDHANADLQAEIN
RLKQELADRIKSLSQGGRASQANPGSTTAKAGQLPSTG

<u>Supplement B:</u> Modules of PARF tandem and optimized nucleotide sequences.

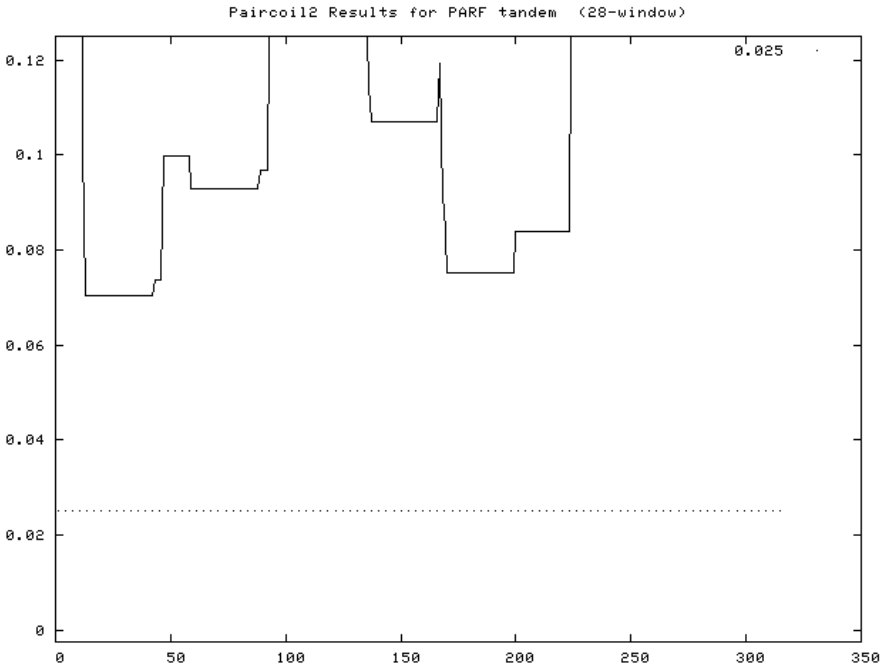
Nucleotide sequence	Amino acid sequence
>LC1903optimized	>LC1903optimized
CTGCAGCAGGCAGAATATCTGCAGCGTCTGAACGATCATTTTGAA	LQQAEYLQRLNDHFE
>stGrobn.0optimized	>stGrobn.0optimized
GATGTGTATACCGAGTATCTGAAACGCCTGAATGATGATGCAGAA	DVYTEYLKRLNDDAE
>stG211.0optimized	>stG211.0optimized
GCAGGTCGTACGGAATATCTGAAATCACTGAACGATTGGGCAGAA	AGRTEYLKSLNDWAE
>stG211.1optimized	>stG211.1optimized
GAACAGAAAGCAGAGTATCTGCGTAGTCTGAATGACTGGGCTGAA	EQKAEYLRSLNDWAE
>stC2sk.0optimized	>stC2sk.0optimized
GATACCCATACCCAATACCTGAAACGGCTGAACGATTATTTCCAA	DTHTQYLKRLNDYFQ
>stG97.0optimized	>stG97.0optimized
GAAGTTTATACGCAGTACCTGAAACGTCTGAATGACCAGTTTCAG	EVYTQYLKRLNDQFQ
>stG4545.0optimized	>stG4545.0optimized
GCAAGCGTTGCAGTTGCACTGACCGTTCTGGGTGCAGGTCTGGCA	ASVAVALTVLGAGLA
>stC6746.0optimized	>stC6746.0optimized
AAAACCTATGCAGATTATCTGAAAACTCTGAATGATCGCTTCGAG	KTYADYLKTLNDRFE
>stCQ343.0optimized	>stCQ343.0optimized
AACACCTACACCGAATACCTGAAATCCCTGAATGATCACTTTCAG	NTYTEYLKSLNDHFQ
>stC46.0optimized	>stC46.0optimized
AAAGATTACGCCAACTATCTGAAAGAGCTGAACGACCACTTTGAA	KDYANYLKELNDHFE
>emm55.0optimized	>emm55.0optimized
AATTGGTATGCAACCTACCTGAAAGAACTGAATGACGAATTTGAG	NWYATYLKELNDEFE
>emm3.0optimized	>emm3.0optimized
CGTCAGAAAGCCGAGTACCTGAAAGGACTGAACGACTGGGCTGAG	RQKAEYLKGLNDWAE
>emm31.5optimized	>emm31.5optimized
ACTCAGAAAGCTGAATATCTGAAAAACCTGAATGACTGGGCAGAG	TQKAEYLKNLNDWAE
>stG10.0optimized	>stG10.0optimized
GGTCAGATGGCAGAGTATCTGAAAGCCCTGAACGATCATTTCGAG	GQMAEYLKALNDHFE
>stG120.1optimized	>stG120.1optimized
GGTGTTCTGGCAAATTACCTGAAAACCCTGAACGACCACTTCGAA	GVLANYLKTLNDHFE
>stG11.0optimized	>stG11.0optimized
GGCGAATATGCACGTTATCTGCAGAAACTGAATGATCAGTTCCAG	GEYARYLQKLNDQFQ
>stG351.0optimized	>stG351.0optimized
GGTGTGTATGCCCGTTACCTGAAAAAACTGAACGATCAATTTCAG	GVYARYLKKLNDQFQ
>stC5344.0optimized	>stC5344.0optimized
GGCGTTTATGCGCGTTATCTGGAAACGCTGAACGACCGTTTTCAG	GVYARYLETLNDRFQ
>stC-NSRT2.0optimized	>stC-NSRT2.0optimized
GGTGTATACGCACGCTATCTGAAACGTCTGAACGATCACTTTCAG	GVYARYLKRLNDHFQ
>stG2574.0optimized	>stG2574.0optimized
AGTGATGATGAAGCATATCTGAAACGGCTGAATGAATGGGCAGAA	SDDEAYLKRLNEWAE

<u>Supplement C:</u> Coiled-coil formation predicted for protein StG11 (C1) and PARF tandem (C2) using Paircoil2 (252). P-score cutoff is 0,025 (dot line). Residues below this score are predicted to be in a coiled-coil. Minimum search window length:28.

C1.



C2.



<u>Supplement C:</u> Known *S. pyogenes* and SDSE *emm*-types that contain a PARF motif.

S. pyogenes	***S. dysgalactie* subsp. *equisimilis***	
emm3	emm31	LC1903
emm31	stG10	stGrobn
emm55	stG120	stG211
	stG11	stC2sk
	stG351	stG97
	stC5344	stG4545
	stC-NSRT2	stC6746
	stG2574	stCQ343
	LC1904	stC46